Department of Applied Physics

Computational Modeling Investigating the Interactions of Proteins with Fullerene-based Nanoparticles

Praveen Nedumpully Govindan

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 12/2013

Computational Modeling Investigating the Interactions of Proteins with Fullerene-based Nanoparticles

Praveen Nedumpully Govindan

A doctoral dissertation completed for the degree of Doctor of Science in Technology to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall E of the school (Otakaari 1F) on the 15th of January 2013 at 12 o'clock.

Aalto University
School of Science
Department of Applied Physics
SOFT

Supervising professor
Prof. Tapio Ala-Nissilä

Thesis advisor
Dr. Emppu Salonen

Preliminary examiners
Prof. Roland Faller, University of California Davis, USA

Dr. Alex Bunker, University of Helsinki, Finland

Opponent
Prof. Peter Tieleman, University of Calgary, Canada

Aalto University publication series
DOCTORAL DISSERTATIONS 12/2013

ISBN 978-952-60-4979-3 (printed)
ISBN 978-952-60-4980-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-4980-9

Unigrafia Oy
Helsinki 2013

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Praveen Nedumpully Govindan

Name of the doctoral dissertation
Computational Modeling Investigating the Interactions of Proteins with Fullerene-based Nanoparticles

Publisher School of Science

Unit Department of Applied Physics

Series Aalto University publication series DOCTORAL DISSERTATIONS 12/2013

Field of research Engineering Physics, Theoretical and Computational Physics

Manuscript submitted 28 September 2012 **Date of the defence** 15 January 2013

Permission to publish granted (date) 21 December 2012 **Language** English

☒ **Monograph** ☐ **Article dissertation (summary + original articles)**

Abstract

In this work, the effects of fullerene-based nanoparticles (fullerene and two of its derivatives) on the structure and functioning of proteins were investigated. An approach combining two computational methods — molecular docking and molecular dynamics simulations — was used. The studies were based on, and complement experiments on the effects on nanoparticles on proteins.

To make a better prospective, first, the properties of nanoparticle clusters in water were studied. The nature and stability of the clusters were found to depend on the surface properties of the nanoparticle. Nanoparticles with hydrophobic surfaces made strong and stable clusters, whereas hydrophilic nanoparticles made loose associations whose structure changed over time. In addition, the effects of nanoparticles on the secondary structure of small peptides were studied. For some peptides, a small increase in the alpha-helix content was observed in the presence of nanoparticles.

For protein-nanoparticle interactions, the inhibition of an enzyme protein, namely *taq* DNA polymerase, by fullerene derivatives was studied. Based on our studies, we predicted that the inhibition was caused by tertiary structural changes of the protein induced by the nanoparticles. Point mutation studies which could be used to examine our predictions were also proposed. In another study, the inhibition of tubulin self-assembly into microtubules by fullerene species was investigated. Simulation studies indicated that binding of nanoparticles to certain locations on tubulin was responsible for the inhibition. These binding sites are important for self-assembly as they are located in areas that make contact with the neighboring tubulins in microtubules. Finally, interactions of FUL and FUOH nanoparticles with ubiquitin was studied. Two more prominent binding sites, including one near the C-terminal tail was observed, and the biological implications of the binding are discussed.

Keywords proteins, nanoparticles, nanotoxicity, molecular dynamics, molecular docking, simulations, computational physics

ISBN (printed) 978-952-60-4979-3 **ISBN (pdf)** 978-952-60-4980-9

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2013

Pages 114 **urn** http://urn.fi/URN:ISBN:978-952-60-4980-9

Preface

The work I present in this Dissertation was carried out in the Computational Soft Matter Group (SOFT), Department of Applied Physics, Aalto University, during 2007–2012. I would like express my deep gratitude to my advisor Dr. Emppu Salonen. I am thankful for the support Emppu gave during all these years, and also for the patience he showed, especially while correcting manuscripts, funding applications and this Dissertation. I also would like to express my sincere thanks to my supervising professor, Prof. Tapio Ala-Nissilä.

This work would not have taken in its current form, had we not collaborated with Prof. Pu-Chun Ke's research group in Clemson University, USA. I am grateful to Pu-Chun for all his encouragements and support. I would like to express my thanks to Dr. Luca Monticelli, INSERM, Paris, for providing me an opportunity to visit his research group, and for a fruitful collaboration. I also would like to thank Prof. Ilpo Vattulainen, Tampere University of Technology. I have really enjoyed working with him and his group.

Before moving to Helsinki, I worked with Prof. P. B. Sunil Kumar in IIT Madras, India. I am grateful to him for his mentoring and support. I also would like to thank Prof. T. S. Natarajan and Prof. Prem Bisht who were my teachers during my years in IIT Madras. At this moment, I remember late Prof. E. K. Narayanan whose love and care have influenced me greatly. I am thankful to Prof. Peter Tieleman for accepting to act as my opponent.

Thanks to National Graduate School in Material Physics, the Helsinki University of Technology and HPC Europa2 for the financial support without which this work would not have materialized. Studies like this one are impractical without super

computing facilities. Thanks to CSC — IT Center for Science, and Triton, the Aalto Science-IT project for the computing time they provided.

I have had warm conversations and delightful discussions with many of my colleagues in SOFT and BIO groups in Finland, and CFL group in IIT Madras. Thanks to each one of you. I am lucky to have a bunch of wonderful friends. Thank you all. Finally, thanks to my family and other relatives for their love and support.

Helsinki, December 21, 2012,

Praveen Nedumpully Govindan

Contents

Statement of contribution

This Dissertation focuses on the interaction between proteins and fullerene-based nanoparticles. Two articles based on the work described in Sections 5.3 and 5.4 have been published:

[I] "Mechanism of *Taq* DNA Polymerase Inhibition by Fullerene Derivatives: Insight from Computer Simulations", P. Nedumpully Govindan, L. Monticelli, E. Salonen, **J. Chem. Phys. B**, 116, 10676 (2012).

[II] "In Vitro Polymerization of Microtubules with a fullerene Derivative", T. A. Ratnikova, P. Nedumpully Govindan, E. Salonen, P. C. Ke, **ACS Nano**, 5, 6306 (2011).

Publication [I] is computational in nature and is based on the work described in 5.3. The author carried out setting up and running the simulations, analyzing the results and writing up the first draft of the article. Publication [II] includes both computational and experimental studies. The computational part is based on the work described in Section 5.4. The author performed setting up and running the simulation, and analyzing the results. He also wrote the first draft of the part concerning simulations, which was then revised by other authors of the manuscript.

A third article based on the results described in Section 5.5 is under preparation and will be published in a peer reviewed journal. Further work is underway for the effects of fullerene nanoparticles on secondary structures described in Section 5.2. The results of this study will also be published in peer reviewed journal. The author has independently carried out almost all parts of this work, including conceiving and planning the research, setting up and running the simulations and analyzing the results.

1. Overview

Since the invention of the computer in the 20th century, computational methods have opened a new paradigm in scientific research. In molecular biology, nowadays they have opened a new window to the length and time scales which are hardly attainable by experimental methods. This Dissertation deals with computational approaches to study biological systems. More specifically, two computational methodologies — molecular dynamics and molecular docking — are used to study the interactions between proteins and fullerene-based nanoparticles.

Any scientific process (see Fig. 1.1) begins by posing a question about a *system*. The systems of interest are often too complex, so to achieve a better understanding, simple *model systems* of the real system are made. A model can be theoretical, computational or even experimental. An experimental model system is one using which an understanding of the important properties of a more complex system can be gained. For example, single component (or multicomponent) phospholipid membranes are often used to understand the properties of the much more complex cell membranes. Computational/theoretical models are the representation of the real system and sets of rules [1]. Simulations, which this work employs, are used to measure quantities using computational models. It complements the theoretical predictions based on model systems and experimental observations of the real or model systems. Comparisons are made between experimental observations, theoretical predictions and computational results, and conclusions are drawn.

As for the systems related to this Dissertation, the scientific question asked, on a big picture, is “what are the effects of nanoparticles on biological systems?”. Since biological systems are too complex, and proteins do most of the important “work” in

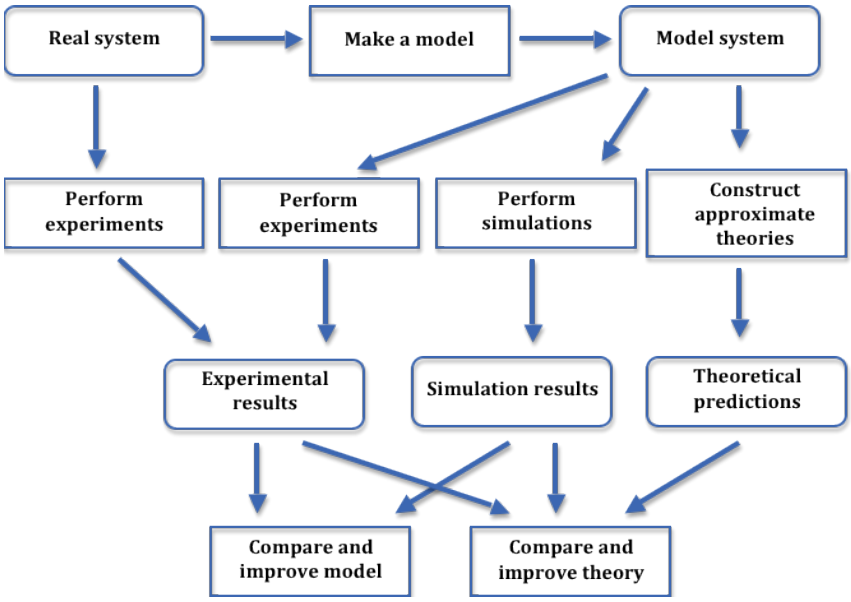


Figure 1.1. A flowchart of the scientific methodology. The work related to this Dissertation is based on computer simulations of model systems.

them, the question is simplified and reframed to "what are the effects of nanoparticles on proteins and their functioning?". As will be discussed in Chapter 3, experiments have shown that nanoparticles indeed affect the structure and functioning of proteins. Now, the next question is "how and why nanoparticles affect proteins?". This Dissertation attempts to answer this question in a few selected cases. In doing so, whenever it is possible, comparisons are made with experimental results.

This Dissertation is organized as follows. In Chapter 2, a general account of proteins, their structure and their functioning are given. The three proteins which are studied in this Dissertation, namely *taq* DNA polymerase, tubulin and ubiquitin, are also described in detail. In Chapter 3, the properties of fullerene-based nanoparticles are discussed. The studies on their effect on biological systems are described briefly. And finally, the experimental results which motivated this work are detailed.

The computational methods used in this work are described in Chapter 4. In Sections 5.3, 5.4 and 5.5 the results on the interaction of nanoparticles with *taq* DNA polymerase, tubulin and ubiquitin are discussed, respectively. The cluster properties of nanoparticles are described in Section 5.1 and their effects on the stabilities of α-helical structures in 5.2. Finally, a brief summary and views on future research directions are given in Chapter 6.

2. Background on proteins

2.1 Introduction

The diversity of life on earth is astonishing. A virus, which is the simplest form of life[1], may consist of only genetic material (DNA or RNA) and a protein coat. But some organisms have trillions of cells with distinct organs performing specific functions (human body, for instance, is a colony of more than 10^{13} cells [2]). Despite this diversity, the fundamental mechanisms by which life progresses in different organisms are strikingly similar. Cells are the structural and functional building blocks of higher forms of life. The basic biological functions of cells are similar in all species. Cells (and some intercellular organelles) are separated and protected from the surroundings by membranes. The genetic information is stored in nucleic acids. Proteins are the "workforce" which perform almost all the cellular functions.

2.2 Proteins

Proteins constitute most of the dry mass of cells [2]. They are called the building blocks of cells because they are responsible for almost all cellular functions [2]. For example, enzyme proteins catalyze chemical reactions necessary for cell functioning. Ion channels and ion pumps are membrane proteins which selectively transport small molecules in and out of cells, thus acting like gates. Proteins such as tubulin help cells to maintain their size and shape by giving mechanical strength. They also

[1] There is also a different viewpoint in which viruses are not life forms but organic structures interacting with life.

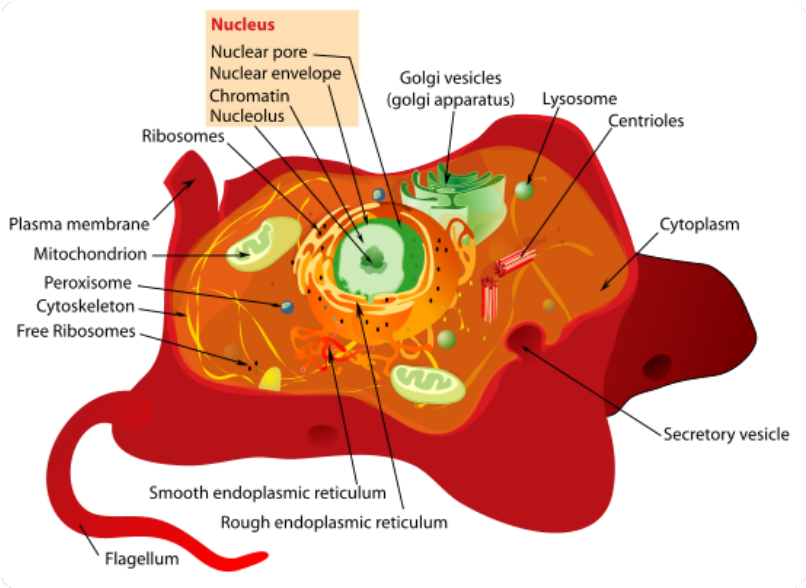


Figure 2.1. A cartoon of the animal cell [3].

play a role in cellular locomotion and cell division. Motor proteins such as kinesin transport cellular organelles and macromolecules. Messenger proteins are involved in transmitting messages between cells. Proteins also function as toxins, antibodies and capsids.

Proteins are co-polymers of their basic subunits called *amino acids* [2]. Typical proteins consist of 50 to 2000 amino acids [2]. Neighboring amino acids in a protein are covalently connected to each other via *peptide bonds* (Fig. 2.2), hence protein chains are also known as polypeptides [2]. Depending on its sequence of amino acids, a protein assumes its unique three dimensional structure [2]. The function of a protein depends on its shape. This is because proteins need to bind specifically to other molecules for their functioning, and this requires a certain surface conformation depending on the binding molecule. Since the shape of a protein is determined by its amino acid sequence, the functioning of a protein also depends on its amino acid sequence.

Amino acids are small molecules which have an amine group and a carboxylic acid group connected to a central *alpha* carbon atom (C_α) [4]. A peptide bond is formed by the chemical reaction between the amine group of an amino acid with the carboxylic group of its neighbor (see Fig. 2.2). The amino and carboxylic groups and

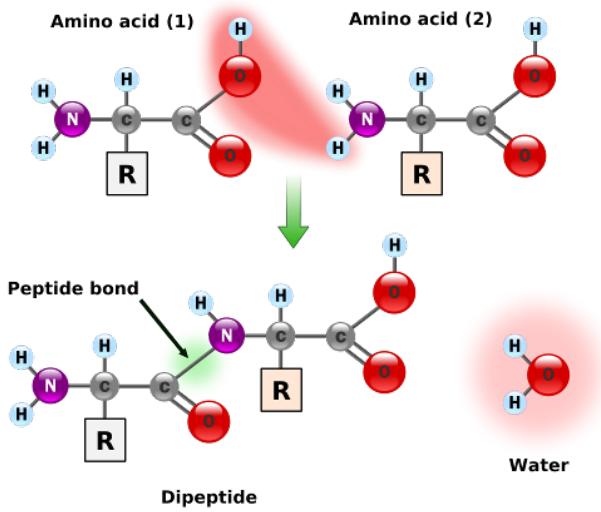


Figure 2.2. Amino acids in a protein chain are connected via peptide bonds. Amino acids have amine and carboxylic groups and a side chain (marked **R**). These three groups are attached to the C_α atom of amino acid [5].

the C_α atom which repeat along a polypeptide chain form the *peptide backbone* or the *main chain* [4]. Another important moiety of the amino acids which does not take part in the peptide bond formation is the *side chain* (marked **R** in Fig. 2.2). Twenty different types of side chains are found in naturally occurring proteins, and hence 20 different types of amino acids [4]. Since the amino acids differ only in the side chains, differences in their nature and properties depend only on the side chains. Some of the side chains are charged whereas some are charge-neutral [4]. Neutral amino acids can be polar or non-polar. One end of the polypeptide chain has a free amine group and the other end has a free carboxyl group. They are called the N-terminus and the C-terminus, respectively. By convention, the amino acids (also called residues) in the polypeptide chain are numbered starting from the N-terminus.

The peptide bond is planar and hence the conformations which a protein chain can assume are greatly limited [4]. The non-covalent interactions between atoms are usually 30-300 times weaker than covalent bonds such as peptide bonds [2]. However, collectively, the non-bonded interactions are significant and they further limit the conformations proteins can assume. Most proteins, under suitable conditions, *fold* into a single free energy minimum conformation called the *native state* [4]. In the native structure, the non-polar residues are usually buried inside the protein and the polar residues are distributed on the surface due to the hydrophobic effect. Other non-covalent interactions which play important role in protein folding are hydrogen

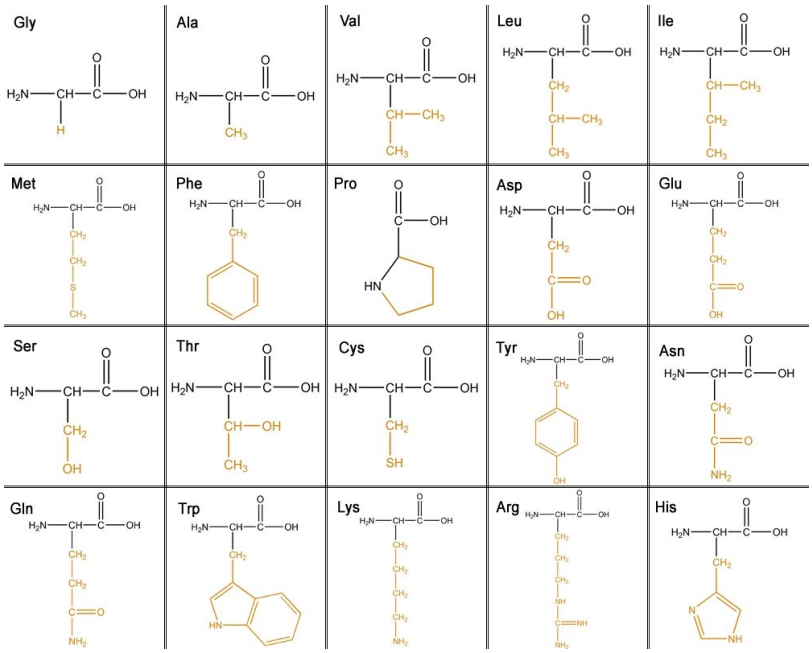


Figure 2.3. Twenty different amino acids are found in naturally occurring proteins [4].

bonds and electrostatic and van der Waals interactions [2].

The native structure can be disturbed by changes in the environment. Proteins lose their native structure or they *denature* by changes in temperature, pH or solvent type [2]. Upon denaturation, proteins lose their functionality, but they usually return to their native states and regain activity if the favorable conditions are restored. Interactions with other molecules can also change the native structure of a protein, which may also affect their functioning. This is a very important aspect for the complex molecular mechanisms by which cells function. Functioning of many proteins is "switched off" or "switched on" by the binding of other molecules. The binding molecules can be small molecules, nucleic acids or proteins themselves. Many drug molecules, for instance, work by binding to a target protein and altering its functioning [6]. As will be discussed later, in this Dissertation the effect of fullerene-based nanoparticles on the structure and functioning of proteins is studied.

Large proteins may have *domains* which fold into their own native structures more or less independently of the rest of the protein [2]. Domains can also function independently. For example, *taq* DNA polymerase, which we will discuss in this Dissertation has three domains that can function independently. Different organisms have

structurally similar proteins to perform a particular task. These proteins may have different amino acid sequences (two proteins may need only about 30% similarity in their amino acid sequences to fold into similar structures). However, if a particular amino acid is seen to be conserved across different species, it is usually an indication for the importance of that amino acid in the functioning of the protein.

2.2.1 Four levels of protein structure

Earlier in this section, we learned that the three-dimensional structure of a protein is very important for its functioning. For ease, the structures of proteins are described at four levels [4] (Fig. 2.4). The amino acid sequence of a protein is called its *primary structure*. The primary structure helps to distinguish proteins with similar structures, and also to identify proteins with similar sequences. The second level is the local *secondary structure* which is discussed in detail in the next section. A full three dimensional structure of a peptide chain is called its *tertiary structure*. The tertiary structure describes the relative positions of all atoms in a peptide chain. If a protein consists of more than one peptide chain (e.g., tubulin dimer), the relative positions and orientations of individual chains describe the *quaternary structure*.

2.2.2 Secondary structure

Although folded structures of different proteins are different, some local patterns are universal. These are called the secondary structures [4]. Secondary structures are formed by the interaction between amino acids close to each other in sequence. The interaction responsible for secondary structures is hydrogen bonding formed between amine and carboxylic groups of the peptide backbone. These bonds do not involve any side chains and hence the universality of secondary structures. However, the side chains of some residues can have an influence on the secondary structures. The two most commonly found secondary structures are *α-helices* and *β-sheets*.

An α-helix is formed when a polypeptide backbone twists around an axis, forming a spring-like structure (Fig.2.5). In principle, both left- and right-handed helices are possible, but left-handed helices are unstable due to steric hindrance. In an α-helix, the hydrogen bonds are formed between the carboxyl group of residue *i* with the

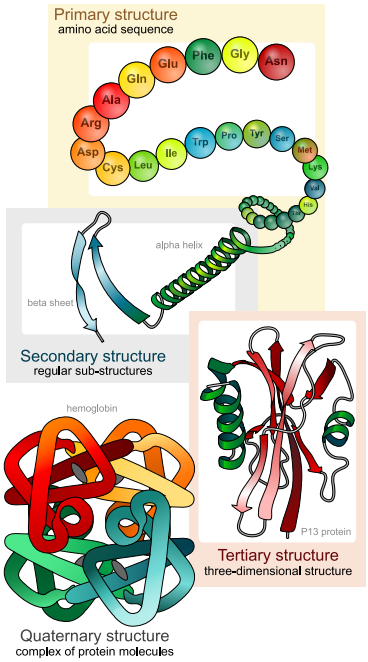


Figure 2.4. The protein structure can be described at four levels as schematically shown here [3].

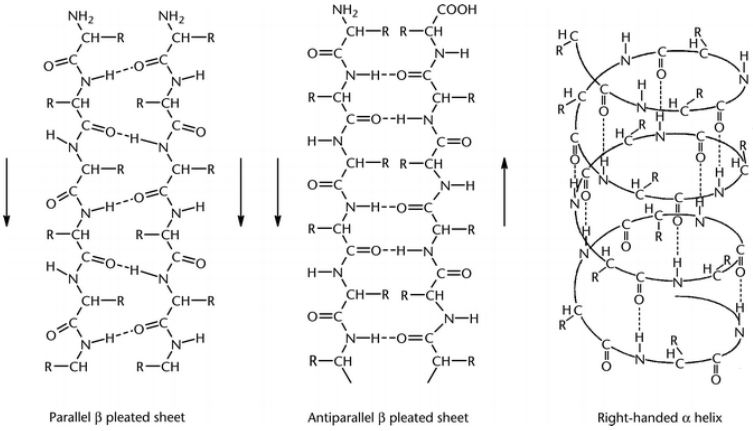


Figure 2.5. The two most common secondary structures of protein: β-sheets (parallel and antiparallel) and α-helix [7].

amine group of residue $i+4$ [8]. A protein is translated by about 3.6 amino acids per turn of an α-helical structure (with a pitch of 0.54 nm) [9]. Some residues, due to their intrinsic nature, are *helix-friendly*, meaning that they favor helix formation. Alanine, leucine and methionine are examples for such residues [10]. Some other residues such as cysteine, aspartic acid and glycine tend to destabilize α-helices [10].

Two less-commonly found helical structures are the 3_{10}-helix and π-helix [8]. In a 3_{10}-helix, the hydrogen bond is formed between residues i and $i+3$, making the structure more compact than the α-helix. In a π-helix, the hydrogen bond formation is between residues i and $i+5$.

β-sheets are formed when hydrogen bonds are formed between two or more peptide strands, called β-strands (Fig. 2.5) [8]. β-strands are typically 3-10 amino acids long. The adjacent strands can be oriented either in the same or in the opposite direction. Thus β-sheets can be parallel, antiparallel or mixed. Residues such as phenylalanine, valine and isoleucine are favorable residues for β-sheet formation [8].

Two other commonly found secondary structures of proteins are turns (also called hairpin bends or β-turns) and loops. Both of them connect secondary structures such as α-helices and β-sheets. The difference between them is that turns are small motifs (typically 3–5 residues) with hydrogen bonds between them, giving them a well-defined structure, whereas loops are longer motifs without a well-defined structure [8]. Turns are seen in regions with a sharp change in orientation such as between antiparallel β-strands. Loops are usually flexible, rich in hydrophilic residues, and are seen on the surface of proteins.

The secondary structure can be determined by looking at the hydrogen bond patterns. An alternative method is by measuring the dihedral angles of the peptide backbone. As mentioned before, the peptide bond is planar and hence non-rotatable, whereas C_α-C and C_α-N bonds of amino acids are rotatable. However, for residues involved in secondary structure formation, rotation about these bonds are limited due to steric hindrance between atoms [4]. Hence, the corresponding dihedral angles can have only certain values. By measuring these dihedral angles, the secondary structure formed can be determined. The dihedral angles around C_α-C and C_α-N bonds are traditionally represented by ψ and ϕ, respectively.

A $\{\psi, \phi\}$ plot for the residues of a protein is called the *Ramachandran plot* [4]. For residues involved in right-handed and left-handed α-helical structures, the favorable $\{\psi, \phi\}$ values are about {-60, -50} and {+60, +60}, respectively [8]. For parallel and anti-parallel β-sheets, the values are {-120, +115} and {-140, +135}, respectively [8]. In experiments, the secondary structure content of a protein is usually determined using circular dichroism (CD) spectroscopy [11]. It is based on the difference in the absorption of left-polarized and right-polarized light by regular structures. α-helical, β-sheet and random coil structures show their characteristic spectrum with far-UV light. Thus by looking at the CD spectrum, it is possible to predict the α-helical and β-sheet contents of a protein [11].

So far in this chapter we have discussed some of the basic aspects about the structure and functioning of proteins. In the following sections we shall discuss three proteins in more detail. They are an enzyme protein called *taq* DNA polymerase, tubulin which is the building block of microtubules, and ubiquitin which is involved in many cellular processes. The functioning of these proteins in the presence of fullerene-based nanoparticles is the theme of this Dissertation.

2.3 *Taq* DNA Polymerase

Taq DNA polymerase belongs to a DNA polymerase family of proteins. The proteins in this family are enzymes involved in the duplication and repair of DNA [12]. The function of DNA polymerases is to catalyze the addition of nucleotides to synthesize new DNA strands [13]. A DNA polymerase uses a *template* DNA strand to determine the nucleotide sequence, and makes a new DNA strand [2]. The short strand to which nucleotides are added one by one, is called the *primer* DNA.

Taq DNA polymerase (*taq* pol in short) is a DNA repair enzyme which adds nucleotides to DNA in the thermophilic bacterium *thermus aquaticus*. A salient feature of *taq* pol is its thermostability [14] — it has the ability to function at very high temperatures due to the thermophilic nature of the bacterium. It is one of the most important enzymes in molecular biology because of its usage in polymerase chain reaction. Polymerase chain reaction (PCR) uses polymerase enzymes to copy DNA *in vitro* (details in section 2.3.3). PCR requires repeated heating and cooling of the

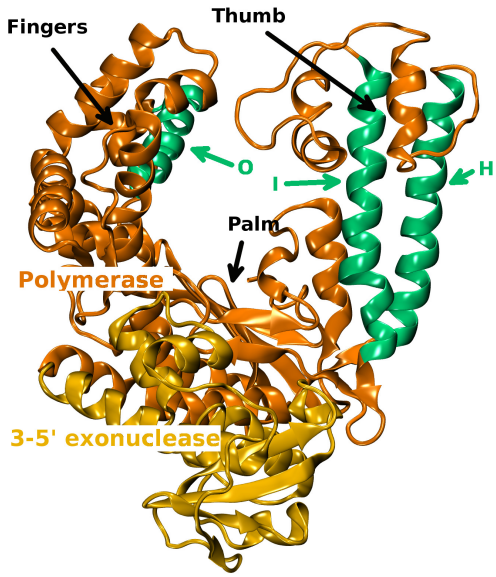


Figure 2.6. Klen*taq* fragment of *taq* DNA polymerase consists of exonuclease and polymerase domains. The polymerase domain has fingers, palm and thumb subdomains [13]. The O, H and I helices of the polymerase domain are highlighted in green.

reaction solution. Because of its thermal stability, *taq* pol is highly preferred in PCR over other polymerase enzymes.

Taq pol has three domains, a 5' nuclease domain, an inactive 3'-5' exonuclease domain and a polymerase domain [15] (Fig. 2.6 shows the exonuclease and polymerase domains, which together form the Klen*taq* fragment). The polymerization reaction is catalyzed by the polymerase domain. This domain looks like a right hand and has three subdomains called *thumb*, *fingers* and *palm* [13] (see Fig. 2.6). Three carboxylic residues in the palm subdomain (Asp 610, Asp 785 and Glu 786) form the active site for the catalysis [15].

An important property of any polymerase protein is its fidelity. Fidelity of a polymerase is the ability to copy a DNA strand without making any mistakes in the nucleotide sequence. *Taq* pol makes one mistake in about 9000 nucleotide additions and is one of the polymerases with low fidelity [16]. This is considered a drawback of *taq* pol and is caused by the inactivity of its exonuclease domain (whose function is to correct errors that the polymerase domain makes). However, *taq* pol is widely used to amplify small fragments of DNA in laboratories. The fidelity of *taq* pol can be improved by cleaving the the 5' nuclease domain, leaving the Klen*taq* fragment

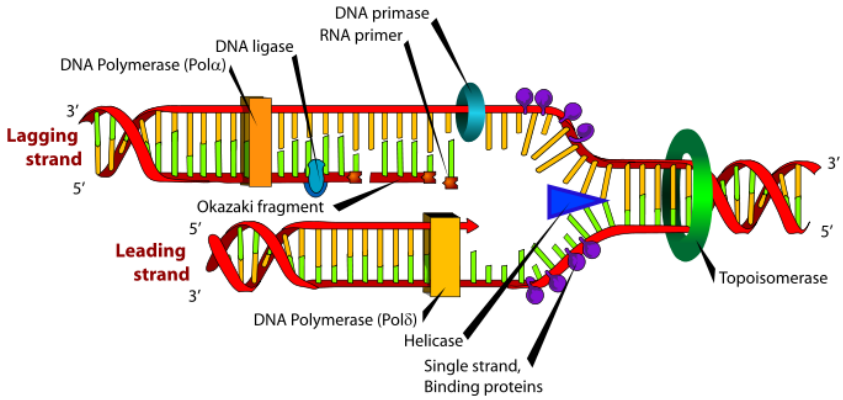


Figure 2.7. The DNA replication fork. In the duplication process, DNA is unwound, strands are separated and each strand is copied by proteins [3].

of the protein [17]. In one of the studies described in this Dissertation, the Klen*taq* fragment is used for simulations although the names *taq* DNA polymerase and *taq* pol are also interchangeably used.

2.3.1 Copying of DNA in cells

DNA consists of two strands which form the famous double helix structure. The sequence of nucleotides in two strands are complementary to each other (due to Watson-Crick base pairing) and anti-parallel. During DNA duplication (see Fig. 2.7), the DNA is unwound by a protein called topoisomerase, and the two strands are separated by another protein called helicase. Two polymerase proteins then bind the two single stranded DNAs, use them as templates, and synthesize two new DNA strands. These two newly synthesized strands complement their templates, so in the end two exact copies of the original double stranded DNA are made.

2.3.2 Functioning of *taq* DNA polymerase

For catalyzing the polymerization reaction, *taq* pol binds to a single-stranded DNA with a bound primer and aligns it properly. During this process, a large conformational change occurs in the thumb subdomain [12]. Provided that a right type of a deoxy-nucleoside triphosphate (dNTP) binds to the protein at the O-helix of the fingers subdomain, the helix undergoes an orientational change [12]. By changing the orientation of the O-helix, *taq* pol makes a transition from an “open” to a “closed”

state. This causes the dNTP to get closer to the DNA and orient in the appropriate way [12]. The polymerization reaction takes place in this closed state [18]. Three acidic residues (Asp785, Glu786 and Asp610) in the palm subdomain form the active site, and catalyze the addition of the nucleoside to the DNA [12]. The fingers subdomain then repositions back to the open state, and the nucleoside addition process is repeated until either the DNA copying is completed or until the DNA unbinds from the protein.

2.3.3 Polymerase Chain Reaction

In laboratories, small fragments of DNA are amplified using PCR. The method was introduced by Kary Mullis in 1983 and uses polymerase enzyme to duplicate DNA [19]. Up to millions of copies of a DNA sample can be produced by repeating the duplication cycle. A PCR cycle involves at least three steps [20]:

(i) Denaturation. The reaction solution containing the DNA sample, short primer DNAs, dNTPs and polymerase is heated above 90 °C. During this step the double stranded DNA is separated into two single strands.

(ii) Annealing. The temperature is lowered to 50–60 °C in this step. Short primer DNAs bind the single stranded DNAs. *Taq* pol binds the template-primer DNA complex.

(iii) Elongation. Nucleotides are added to the primer during this step. The polymerase enzyme catalyzes the nucleotide addition. The temperature depends on the optimal temperature of the polymerase. For *taq* pol the temperature is typically 75–80 °C

Many proteins denature if heated to above 90 °C. However, *taq* DNA polymerase is thermophilic and does not denature or lose activity at high temperature and hence it is routinely used in PCR.

2.4 Tubulin and microtubule assembly

As mentioned earlier in this Chapter, some proteins are involved in giving mechanical strength to cells and cellular organelles. The cytoskeleton is an important cellular component whose main purpose is to give mechanical strength [2]. It is also necessary for many cellular processes such as inter-cellular transport, cell movement

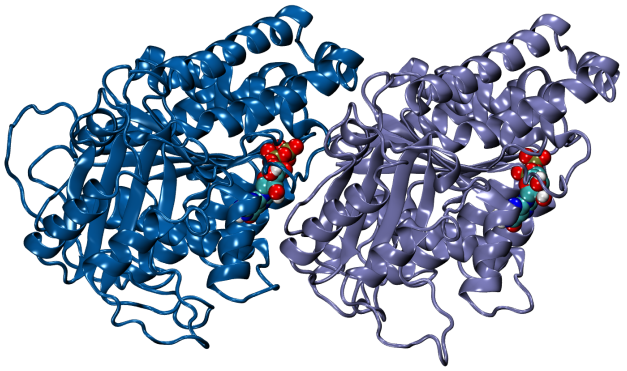


Figure 2.8. Tubulin dimer is made of α- and β-tubulins. GTP (left) and GDP (right) molecules are bound to the N- and E-sites, respectively. The atomic coordinates of the structure were obtained from Protein Data Bank (PDB ID: 1TUB [24]), and the figure was created using the molecular graphics software VMD [25].

and cell division [21]. The eukaryotic cytoskeleton has three constituents: microfilaments, intermediate filaments and microtubules. Of these, microtubules are cylindrical structures with a diameter of about 25 nm [22]. They are highly dynamic structures and grow up to many micrometers in length [23].

The basic building blocks of microtubules are tubulins [2]. Each tubulin, in turn, is a dimer of two structurally similar globular proteins called α- and β-tubulins [26], bonded non-covalently (Fig. 2.8). Both α- and β-tubulins have one guanosine triphosphate (GTP) binding site each. The GTP binding site of α-tubulin is at the α–β interface of the dimer. The GTP molecule bound at this site is stable and non-exchangeable, hence the site is called the non-exchangeable site (or N-site) [2]. The GTP bound to β-tubulin, on the other hand, can be hydrolyzed to guanosine diphosphate (GDP). The resulting GDP molecule at this site can be exchanged with a GTP, hence the binding site is called the exchangeable site (E-site) [2].

Polymers formed by the non-covalent addition of tubulin dimers are called the *protofilaments* [2]. A hollow tube-like structure or a microtubule is formed by the bundling of 13 protofilaments with lateral interactions [2]. Since dimers can be added only head-to-tail to the protofilaments, the two ends grow differently. The growing end will have only β-tubulins exposed while the other end has only α-tubulins exposed. The two ends are called the **+** end and **-** end, respectively. The microtubules are thus said to have a *polarity*.

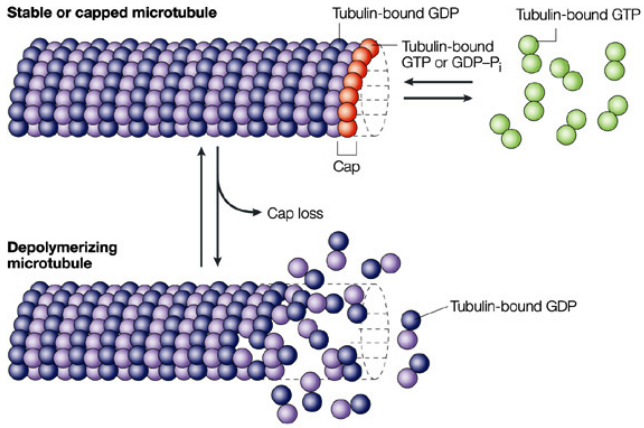


Figure 2.9. Microtubule assembly and disassembly [27]. GTP-bound tubulins (shown in green in solution, and red in microtubule) are assembled into microtubules, which can then hydrolyze the bound GTP molecule to a GDP molecule and a phosphate. The GDP-bound tubulins (blue) fall off microtubule in the absence of a GTP-bound tubulin cap (red).

Another important feature of the microtubule is its *dynamic instability* [28]. It is the phenomenon because of which microtubules switch between growth and shrinkage phases. To understand dynamic instability, one has to know how microtubule assembly takes place. As explained earlier, the N-site of a tubulin always binds a GTP molecule whereas the E-site can bind either a GTP or a GDP molecule. Depending on which one of them binds to the E-site, a tubulin is called to be either in a GDP-bound or GTP-bound state. GTP-bound tubulins have straight contact between their monomers, whereas GDP-bound tubulins are structurally different with a kink between the monomers [29]. To be able to take part in the microtubule assembly, a tubulin has to be in the GTP-bound state [2], a GDP-bound tubulin cannot take part in the microtubule assembly [30].

The GTP-bound tubulins in a microtubule can hydrolyze GTP to GDP and a free phosphate [31]. This process is stochastic in nature. The resulting GDP-bound tubulin has a tendency to disassemble from the microtubule. However, whether the GDP-bound tubulin falls off or not depends on its location. If it is located at the + end of the microtubule, it falls off immediately. However, if it is not located at the + end, it is interlocked between other tubulins (which may be GTP- or GDP-bound) [32]. One has to keep in mind that free GTP-bound tubulins, if available above certain concentration, are constantly added to the + end. This means most of the time the + end grows slowly, leaving behind tubulins in GTP- or GDP-bound states. If tubulins at

the + end hydrolyze their GTP to GDP, they and all the GDP-bound tubulins behind them fall off [31]. This is called a shrinkage phase or catastrophe. This catastrophe continues until a GTP-bound tubulin is reached. In a catastrophe the length of a microtubule decreases rapidly.

2.5 Ubiquitin

Ubiquitin is a small protein consisting of 76 amino acids [4]. It is found ubiquitously (hence the name) in all eukaryotic cells, either as free or attached to other proteins. It is highly conserved [4], meaning that its amino acid sequence remains almost unchanged across families of organisms. For example, 73 out of the 76 amino acids in human and *Arabidopsis thaliana* (a plant) ubiquitins are the same.

The most important function of ubiquitin is related to protein degradation [2]. Degradation of unwanted proteins is important for the cellular functioning, and is carried out by the ubiquitin-proteasome system. The role of ubiquitin in protein degradation was discovered in 1970 and the discovery was awarded the Nobel prize in Chemistry in 2004 [33]. Chemical attachment of ubiquitin to target proteins serves like a label for their degradation [34]. Proteins labeled for degradation are broken down to smaller peptide chains (a few amino acids long) by a protein complex called proteasome [34], and the peptides are reused for the synthesis of new proteins.

The chemical attachment of ubiquitin to a protein requires three associated proteins [4]. The first one, called the ubiquitin-activating enzyme or E1, activates ubiquitin as its name suggests [4]. The activation happens by the attachment of terminal carboxylate group of ubiquitin to E1 by a thioester bond. In the second step, ubiquitin is transferred to a ubiquitin-conjugating enzyme E2 [4]. Finally, the attachment of ubiquitin to the actual target protein is catalyzed by a ubiquitin-ligase enzyme E3 [4]. E1, E2 and E3 are not specific proteins; a number of different proteins serve as E1, E2 or E3. Attachment of a single ubiquitin to a target protein does not necessarily mark it for degradation, but attachment of a chain of four or more ubiquitins does. The degrading proteasome complex has a catalyzing subunit and a regulatory subunit which recognizes polyubiquitin chains.

Protein degradation is not the only cellular process ubiquitin is involved in. Ubiqui-

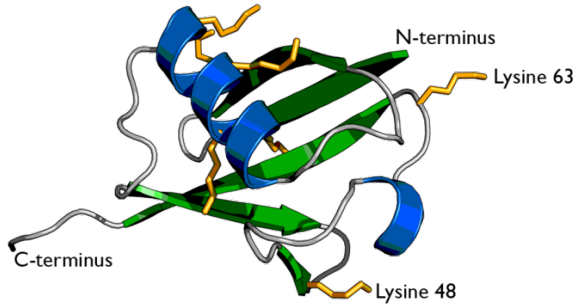


Figure 2.10. Cartoon representation of a ubiquitin molecule. In polyubiquitin chains, the C-terminal residue of one ubiquitin molecule is attached to Lys48 or Lys63 residues of the subsequent ubiquitin [35].

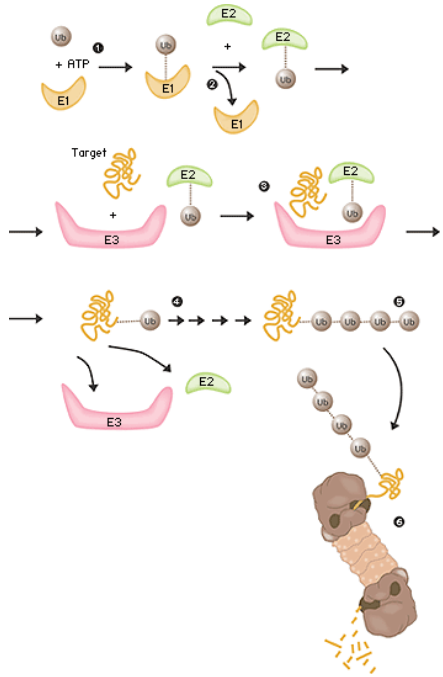


Figure 2.11. Degradation of a target protein by the ubiquitin-proteasome system. The attachment of ubiquitin to a target protein requires three enzymes (E1, E2 and E3). The ubiquitinated protein is broken down by proteasome [33].

tin also takes part in protein trafficking, DNA repair, the assembly of protein signaling complexes, immune function and the activation or inactivation of enzymes [36]. Whether or not proteins are marked for destruction by ubiquitination depends on how the individual ubiquitins are connected to each other in the chain [2]. If the linking is through Lys48 of one ubiquitin and the C-terminal glycine of the consecutive ubiquitin, the target protein is degraded by proteasome [36]. Chains linked through Lys63 and C-terminal glycine are involved in protein trafficking or signaling [36].

Given the role of ubiquitin in many cellular processes, it is not surprising that any failure of ubiquitin-associated processes would cause diseases and disorders. These include cancer, neurodegenerative disorders and dysfunctioning of the immune system. For example, human papilloma virus (HPV) causes the activation of a specific E3 enzyme. This, in turn, leads to the destruction of proteins that control DNA repair, which can result in cancer [4]. In fact, 5.2% of all cancers are caused by HPV infection [37].

3. Fullerene nanoparticles and their biological effects

3.1 Nanoparticles

A nanoparticle, by definition, is a particle whose size is smaller than about 100 nm at least in one dimension [38]. Physical and chemical properties of nano-sized materials are different from their macroscopic counterparts. The properties of nanoparticles are not only different from those of their bulk counterparts but also depend on the size, shape, surface properties of the nanoparticle [39]. For example, the color of gold and silver nanoparticles changes as their size is changed [40]. Similarly, reaction rates of particles increase as the size decreases because of the increase in total surface area. Solubility changes dramatically with changes in the surface chemistry [41]. The toxicity of fullerene-based nanoparticles changes over 7 orders of magnitude by changing the surface derivatization [42]. Since in this fashion it is possible to alter the properties of nanoparticles according to the need, they are promising for many applications [43, 44].

In the early 1990s new and efficient methods were developed to synthesize and manipulate nanoparticles. Since then nanoparticles have been the subject of research in physics, chemistry, materials science and even biomedicine [45–47]. Nowadays it is possible to tune the size, composition and surface properties of certain nanoparticles. As a result, their application and usage increased tremendously, both in industrial products and for research purposes. There are more than 1300 nanoparticle-based products already in the market, and this number is likely to increase [48]. Nanoparticles are thus nowadays produced in large quantities and their entry to the environment

is inevitable. Concerns about their environmental and biological effects are growing as more is understood about their fate in the environment [39, 49].

Some of the nanoparticles released to the environment are able to reach inside organisms through inhalation, ingestion or through contact with the skin [50]. Fine carbon nanoparticles are reported to get deposited in lungs through inhalation [51]. Inside organisms they are able to cross various biological barriers. They are even able to cross the blood-brain barrier — one of the most difficult biological protective barrier to cross — and to reach the brain [52]. Nanoparticles such as fullerenes are able to partition into membranes surrounding the cells. They have also been reported to reach inside the cells and to inter-cellular organelles [53]. Inside cells (and also on the surface of cells) they can interact with a vast number of biomolecules, affect their functioning and hence the functioning of the entire cellular machinery.

The effect of nanoparticles on cellular machinery can be boon or bane and cannot be predicted *a priori*; studies have reported both their beneficial and harmful effects. Nanoparticles show antibacterial [54–58] and antiviral [59–61] activities, reduce the growth of tumor cells [62] and act as antioxidants [63, 64], and function as drug carriers [61], to name a few beneficial effects. Adverse effects include acting as cytotoxic substances [42, 65, 66], inducing cell contractions [67], affecting the functioning of proteins and nucleic acids (see section 3.3 for more details) and inducing oxidative stress to cells [68]. Although concerns of harmful effects of nanoparticles on biological systems are premature, it is important to address them for the healthy development of nanotechnology [39]. In order to minimize risks and maximize the benefits of nanoparticles, more research at the fundamental level is needed [39]. Development of mathematical and computational models has been identified as an area of priority that needs to be addressed in order to develop safe nano-medicines [69].

As discussed in the previous Chapter, proteins play a very important role in cellular machinery by complex interactions between themselves and with other types of molecules. If one or more of these interactions are altered by nanoparticles, the cellular functioning can be disrupted. Thus to understand how xenobiotic substances affect cellular functioning, knowledge of their effect on important proteins (and other molecules) is necessary.

The nanoparticles whose biological effects are studied here are the fullerene C_{60}, also

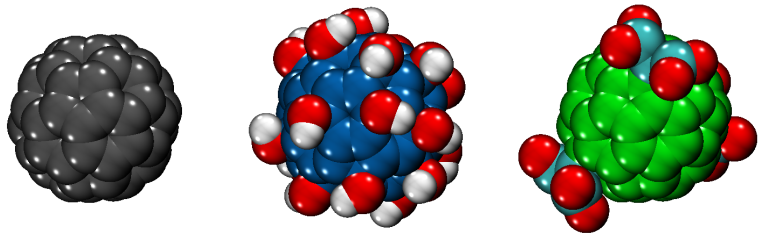


Figure 3.1. Fullerene and its derivatives studied here. Pristine fullerene C_{60} (left), fullerol $C_{60}(OH)_{20}$ (middle) and fullerene trimalonic acid $C_{63}(COO)_6$ (right).

known as the buckminsterfullerene, and two of its derivatives, fullerol $C_{60}(OH)_{20}$ and fullerene trimalonic acid $C_{63}(COO)_6$ (Fig. 3.1). In the following section, properties of these three nanoparticles are discussed and in Section 3.3, previous findings on their effects on biosystems are described. In Section 3.4, the motivation for this work is discussed. The aim of this work is to understand the molecular mechanisms by which these nanoparticles affect the functioning of proteins.

3.2 Fullerene and its derivatives

Fullerene is the third allotrope of carbon discovered, after diamond and graphite. Fullerenes are caged structures formed of pentagons and hexagons of carbon atoms. Unlike diamond and graphite, which can be “infinite”, fullerenes are finite; the number of carbon atoms in a fullerene is fixed. Thermodynamically, larger planar structures are stable but if the number of atoms is limited, the surface curves and forms a caged fullerene. The atoms are sp^2 hybridized and each one is bonded to three other carbon atoms. Fullerenes consist of many hexagons and pentagons. The bonds shared between two hexagons (6:6 bonds) have a double bond nature, and are shorter than the other bonds (6:5 bonds, shared between one hexagon and one pentagon).

The term fullerene represents many molecules with spherical or ellipsoidal shape. Carbon nanotubes and graphite are closely related to fullerenes [70]. Fullerene was discovered in 1985 and nanotubes in 1991. Fullerenes occur naturally [71] and are found even in outer space. C_{60} fullerene (Fig. 3.1), which is made of 60 carbon atoms, is probably the most commonly used fullerene for research purposes, C_{70} fullerene with 70 carbon atoms being the second one. C_{60} fullerene is spherical with

a diameter of about 0.7 nm [72, 73].

Solubility

One of the key issues with the use of fullerenes is that they are virtually insoluble in water due to their hydrophobicity. The solubility of C_{60} fullerene is less than 10^{-9} mg/L in water [74], but it is soluble in many other solvents such as hydrocarbons [73, 75]. To solvate fullerene in water, different techniques can be used [41] which include (a) encapsulation in special carriers, (b) use of co-solvents to suspend fullerenes, (c) chemical modification of fullerene surface and (d) suspending with the help of sonication. C_{60} fullerenes are effectively dissolved in water by encapsulating them in compounds such as γ-cyclodextrin and calixerene [76] or in micelles [77, 78] and liposomes [77]. Fullerenes are also dissolved in detergents or coated with molecules to improve solubility. Polyvinylpyrrolidone (PVP) [79], gallic acid [67, 80] and natural organic matter (NOM[1]) [81] are examples of such compounds. Among these, NOM is particularly interesting because of its presence in the environment. When nanoparticles are released to the environment, they come in contact with NOM and stable complexes may be formed [82, 83].

Without encapsulating fullerenes, their solubility can be increased dramatically by attaching hydrophilic groups to their surface. Polyhydroxylated fullerenes (called fullerols or fullerenols, see Fig. 3.1), for instance, are very soluble in water even though their structure is not very well defined [41]. Lamparth and Hirsch synthesized a highly soluble malonic acid fullerene derivative (fullerene trimalonic acid, see Fig. 3.1) with a well defined three-dimensional structure [84]. Solubility can also be enhanced by attaching water-soluble polymers [62], amino acids [85] and many other hydrophilic moieties to the fullerene cage.

Cluster properties

Dry C_{60} fullerene lump (fullerite) forms the simple cubic structure at low temperatures and the face-centered cubic (fcc) structure with a lattice constant of 1.417 nm above 249 K [86]. In water, due to their hydrophobicity, C_{60} molecules cluster and

[1]NOM is a collection of broken down matter from dead organisms. It consists of proteins, lipids, sugars and many chemicals.

form colloids called nano-C_{60}. The sizes of nano-C_{60} vary from tens to a few hundreds of nm [56, 68]. In vacuum, fullerenes form clusters. Clusters of 13 and 55 fullerenes are observed to form particularly stable icosahedral structures [87].

Production

To produce fullerenes, graphite rods are evaporated by passing a large electric current through them. The resulting fullerene soot contains a mixture of C_{60}, C_{70}, nanotubes and many other nanoparticles. The fullerenes are then extracted and separated [73]. In order to synthesize fullerols, multiple hydroxyl groups are introduced on the fullerene cage by treating it with sulfuric acid [88], sodium hydroxide in the presence of a catalyst (tetrabutylammonium hydroxide) [89] or a mixture of hydrogen peroxide and sodium hydroxide [90]. Fullerene malonic acid derivatives are synthesized by first producing an adduct of fullerene with diethyl bromomalnate, which is converted to the malonic derivative by hydrogenolysis [84, 91].

3.3 Effect of fullerene nanoparticles on biosystems

Effects of fullerene nanoparticles on biosystems have been extensively studied. Both beneficial and harmful effects have been reported. In the following, some of these studies are discussed.

Cell/Organism level

C_{70} fullerenes and $C_{60}(OH)_{20}$ fullerols induce cell lysis [92] and cell contraction [67], reducing the volume of the cells by about 50%. Fullerene derivatives were reported to cause toxicity to human cells [42]. The extent of toxicity depends on the surface functionalization of fullerene [42]. Fullerene derivatives accumulated preferentially in tumor cells compared to healthy cells and showed anti-tumor activity with no apparent side effects [62, 93]. An anti-fullerene antibody was produced in mice [94]. Antiapoptosis activities of fullerene trimalonic acid derivative were reported by Huang *et al.* [95]. Inhaled carbon nanoparticles were able to cross the blood-brain barrier to reach brain in rats [52]. In fish fullerene nanoparticles induced oxidative stress in brain [68]. The mortal effects of C_{60} cluster on two aquatic

species depended on the procedure with which the fullerenes were dissolved in water. Nano-C_{60} dissolved using tetrahydrofuran was mortal whereas water-stirred nanoC_{60} showed much less toxicity [96].

Membranes

Fullerenes are reported to be able to cross cell membranes and reach inside cells [53]. Fullerenes cause damage to membranes by inducing lipid peroxidation [97]. The damage is reported to be caused by reactive oxygen species (ROS) [98]. Lipid composition, phase transition temperature and fluidity of cell membranes are altered by nano-C_{60} [99]. MD simulation studies of fullerene and fullerol with lipid membranes showed that fullerenes translocate into the membrane and stay at about 1 nm from the center [100], while the preferred location of fullerol is near the head group, about 2.3 nm from the center. Coarse grained MD simulations of fullerenes showed that fullerenes prefer to aggregate while in water (forming nano-C_{60}), and it is favorable for the aggregates to partition into the membrane than staying in water [101]. Further, the aggregates tend to disaggregate once they reach the hydrophobic tail region of the membrane. Similar insertion of small carbon nanotubes into membranes has also been observed in MD simulations [102].

Nucleic acids

In the presence of light, fullerene and its derivatives are capable of cleaving DNA molecules [103–106]. In 1994 Nakamura and coworkers reported guanine-specific cleavage of DNA by a fullerene derivative [103]. The cleavage was believed to be induced by reactive oxygen species, which include singlet oxygen, superoxide radical ion and hydroxyl ion, produced by fullerene [105]. Fullerenes are also capable of distorting the DNA structure. MD simulations by Zhao and others showed that C_{60} binds and deforms A-form double stranded DNA and single stranded DNA, but does not change the overall shape of B-form double stranded DNA [107]. The binding energies were in the range from -113 to -176 kJ/mol. Further, simulations of C_{60} derivatives with single stranded DNA showed that the interaction energy and interaction pattern depends on the surface modification of the nanoparticle [108]. Strong interaction with the DNA backbone was seen in simulations with fullerol [109].

Proteins and enzymes

Fullerene forms aggregates with lysozyme [110] and fullerol reduces the ability of lysozyme to dissolve bacterial cell wall by up to 50% at 1:5 molar ratio (lysozyme:fullerol) [111]. The secondary structure of the protein was also altered by fullerol: the α-helical content was reduced from 41.8% to 29.1% at 1:1 molar ratio [111]. Fullerene molecules form water soluble complexes with bovine serum albumin (BSA) without any significant change in the secondary structure of the protein [112]. Similarly complex formation of human serum albumin (HSA) with the malonic acid derivative of fullerene has also been also reported [113].

The effect of 8 fullerene derivatives (including one dimalonic acid derivative) on the functioning of human immunodeficiency virus reverse transcripatase (HIV-RT) and hepatitis C RNA-dependent RNA polymerase (HVC-RP) were studied by Mashino *et al.* [114]. They found that nanoparticles inhibit the functioning of the enzymes. The IC_{50} value[2] for dimalonic acid derivative was found to be 1.2 μM and 3.2 μM for HIV-RT and HVC-RP, respectively. A substrate-dependent inhibition of the activities of rodent liver Glutathione-S-transferase (GST) by polyvinylpyrrolidone-solvated fullerene was reported by Iwata *et al.* [115]. Fullerenes also inhibited human GST and moderately inhibited glutathione peroxidase and glutathione reductase [115]. Meng *et al.* reported that fullerol and TMA derivatives inhibit Moloney Murine Leukemia virus reverse transcriptase [116].

Based on computational studies, Friedman and coworkers predicted that fullerene derivatives fit the hydrophobic cavity of HIV1-Protease (HIV1-P) and hence can be an inhibitor [60]. A fullerene derivative indeed reduced the HIV1 infection of blood cells with an EC_{50} value 7 μM, with no toxic effect on uninfected cells [60]. Tuckerman and coworkers used MD simulations to find that a fullerene core in the hydrophobic core of HIV-P affects the opening of protease flaps [117]. The presence of fullerene causes a barrier for flap opening and the location of the barrier depends on the protonation state of catalytic Asp residue oxygens. They further discovered that the cavity undergoes structural changes for accommodating fullerene core more effectively and consequently more water molecules are excluded from the region.

[2] IC_{50} is the concentration of an inhibitor at which 50% of the maximal activity is lost. EC_{50} is the plasma concentration of the inhibitor at which 50% of the activity is lost in *in vivo* trials.

Enzyme	Nanoparticle	Notes	Ref
MuLV RT	FUOH, TMA	IC_{50}= 89 μM and 39 μM	[116]
HIV RT	DMA, 7 others	IC_{50}= 1.2 μM (DMA)	[114]
HIV P	Many C_{60} derivatives	K_i= ~μM to ~100 nM	[60, 119]
Lysozyme	FUOH	50 % activity at 1:5 lys:FUOH	[111]
HVC RP	DMA, 7 others	IC_{50}= 3.2 μM (DMA)	[114]
Taq pol	TMA, FUOH	IC_{50}= 6.0 μM (TMA)	[120, 121]
GST	FUL	Substrate-dependent inhibition	[115]
Hind III	TMA	IC_{50}= 16.3 μM	[122]
EcoR I	TMA	IC_{50}= 6.0 μM	[122]

Table 3.1. Enzymes inhibited by fullerene and its derivatives. DMA = Dimalonic acid C_{60} derivative.

The effect of fullerene on bacterial potassium channels (KcsA and MthK) and voltage gated potassium channel (Kv1.2) were studied by Kraszewski *et al.* using computational methods [118]. They found that fullerene does not bind or block the selectivity filter region of the channel, but binds to the hydrophobic residues of the extra-cellular loops. If placed in the intracellular region, fullerene binds to the internal cavity entrance and causes a conformational change to a helix which would lead to a change in ion conduction pathway [118].

3.4 Motivation for this work

The studies described in this Dissertation are motivated by many experiments, especially those conducted in Prof. Pu-Chun Ke's research group in Clemson University, USA. These experiments deal with the effect of fullerene-based nanoparticles on the functioning of proteins.

Effects on polymerase chain reaction

Amplification of a heat shock transcription factor (HSTF) gene by polymerase chain reaction was studied in the presence of fullerol [109]. In these experiments, as the concentration of fullerol was increased from 0.2×10^{-4} mM to 1.0×10^{-4} mM, the amount of amplified double stranded DNA was reduced and no double stranded DNA was observed at concentrations higher than 1.6×10^{-4} mM. The hydrodynamic diameter was measured using dynamic light scattering and the values obtained for protein

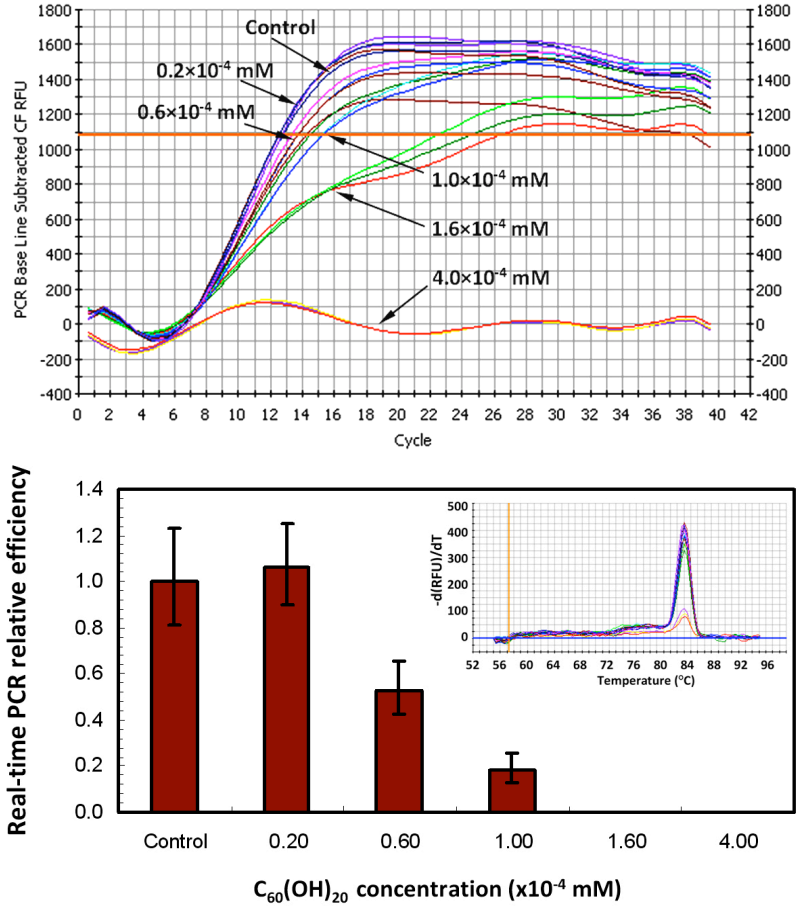


Figure 3.2. Real time PCR efficiency at different fullerol concentrations [109].

and fullerol were 10.8 ($\pm$1.9) nm and 1.0 ($\pm$0.3) nm, respectively, whereas the hydrodynamic diameter of the mixture was 21.9 ($\pm$10.9) nm. The reaction mixture with 4.0×10^{-4} mM fullerol showed the highest UV absorbance. Since double stranded DNA absorbs about 20% less than ssDNA and up to 40% less than dNTPs of the same quantity, the higher absorbance at 4.0×10^{-4} mM fullerol further supports the PCR inhibition.

The inhibition could be a result of the interaction of fullerols with the enzyme, single stranded DNA or with dNTPs in the solution. At the highest fullerol concentration of 4.0×10^{-4} mM, the fullerol to dNTP ratio was 1:500. Thus, most of the dNTPs were free and it was unlikely that the inhibition was caused by fullerol-dNTP binding. The amount of *taq* polymerase present in the reaction solution was only given in enzyme

units (U)[3], so the exact concentration of the protein was not known. But when the amount of *taq* polymerase was increased from 1 U to 4 U, the amplification was partially restored.

In another study Meng *et al*. studied the effect of fullerol and trimalonic acid derivative on the polymerization of β-actin cDNA [121]. As the concentration of nanoparticles was increased from 0.005 mM to 1.0 mM, the PCR products decreased. The PCR was completely inhibited (no product was visible) at concentrations above 0.01 mM and 0.005 mM for fullerol and trimalonic derivative, respectively. To understand which interaction causes the inhibition, the concentrations of the template DNA and enzyme were changed. As amount of template DNA was increased by a factor of four, little change in the product was seen in the absence or presence of fullerol. On the other hand, when the amount of *taq* pol was increased from 3.8 U to 15.2 U, the inhibition was reversed. At 0.01 mM fullerol concentration, the product reappeared at enzyme concentration of 7.6 U, and for the same amount of trimalonic derivative, 11.4 U of enzyme was needed to recover the reaction. Thus the authors concluded that the cause of inhibition was the interaction between the nanoparticles and the enzyme [121]. Further, trimalonic derivative was found to have stronger inhibition activity compared to fullerol.

From these studies, it was concluded that fullerol and trimalonic derivative are capable of inhibiting the polymerization activity of *taq* DNA polymerase. Since no change in the reaction was observed as the template DNA was increased, the inhibition is not likely to be competitive. That is, it is unlikely that the nanoparticles inhibit by binding only to the DNA binding site.

Effects on microtubule self assembly

The effect of fullerol on the polymerization of tubulins was studied by Ratnikova *et al.* [120]. In control experiments without any fullerols, tubulin dimers self-assembled to form microtubules. The synthesized microtubules had a bimodal distribution with peaks at 5 and 20 μm (see Fig. 3.3). When fullerols were added to the reaction mixture, the microtubule self-assembly was inhibited. As the concentration of fullerols

[3]One enzyme unit is the amount of enzyme that catalyzes its substrate at a rate 1 μmole/min under specified conditions.

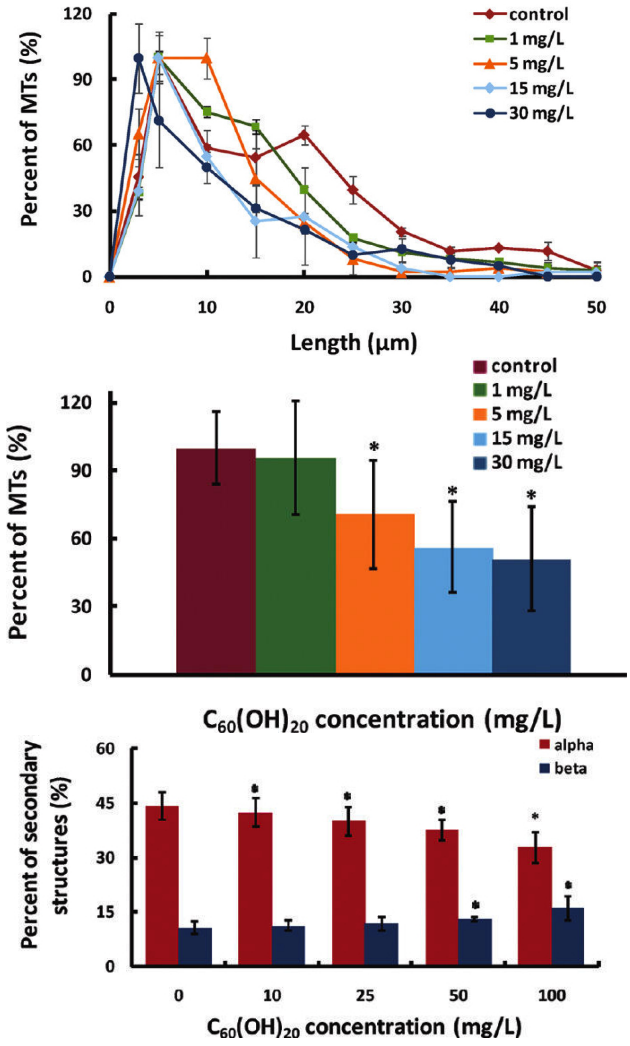


Figure 3.3. Effect of fullerol on microtubule self-assembly. Microtubule length distribution (top), number of microtubules (middle) and secondary structure content at different concentrations of fullerol [120].

increased from 1 mg/L to 30 mg/L (1 mg/L = 5.45 μM), the distribution became a single-peak one centered at 3 μm. This meant that the self-assembly of tubulins into microtubules was greatly affected in the presence of fullerols.

The effect of fullerol on the secondary structure of tubulin was studied by measuring the circular dichroism (CD) spectrum of the samples (Fig. 3.3). The CD spectrum was altered by the presence of fullerol, suggesting conformational change of the protein. As the fullerol concentration was increased, the number of α-helical structures was decreased and the number of β-sheet structures increased (Fig. 3.3). The total number of helical and sheet structures was decreased by 6.9%.

Protein-nanoparticle corona and association of fullerene nanoparticles with ubiquitin

When nanoparticles enter inside an organism, they encounter a huge number of biomolecules, especially proteins. Human serum, for instance, has at least 375 distinct proteins [123]. As a result, soon after their entry, large nanoparticles or large clusters of smaller nanoparticles are coated with one or more layers of proteins, forming a structure known as "protein corona" [124]. Since cells "see" only the protein layers, and not nanoparticles, the fate of the nanoparticles in living organism is determined by the proteins which cover them. The size, curvature, hydrophobicity and chemical properties of the nanoparticle determine which proteins get adsorbed on to them [124]. The composition of the corona changes over time, but the total amount of proteins remains the same.

Ubiquitin is a very common protein found almost in all tissues and also in blood. In the context of nanoparticle-protein corona, the association between fullerol and ubiquitin was studied [125]. Fullerol suspensions with size distributions peaking at 1.5±0.2 nm and 5.0±0.7 nm were mixed with ubiquitin. At 1:1 ubiquitin:fullerol (U:F) molar ratio, the hydrodynamic size of the mixture was measured as 6.0±0.8 nm and 9.7±1.3 nm. At lower (U:F ratio 5:1) and higher (U:F ratio 1:5) concentrations of fullerol, the hydrodynamic sizes were 6.3±1.0 nm and 7.5±0.8 nm, respectively. The zeta potentials[4] of the protein, fullerol and the mixture were also measured for

[4]Zeta potential is an indicator of the stability of colloidal particles or macromolecules against aggregation. When charged particles or macromolecules are placed in a liquid, they are covered by a layer of immobile ions, and a second layer of mobile ions. Zeta potential is

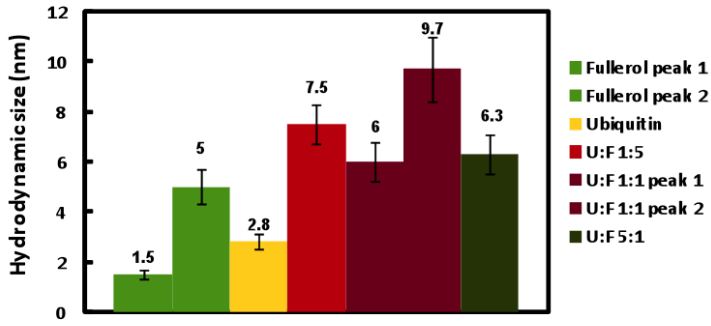


Figure 3.4. The hydrodynamic sizes of fullerol, ubiquitin and their complexes at different molar ratios.

different U:F ratios. For fullerol without ubiquitin, the zeta potential was large and negative at -58.4±6.5 mV and for ubiquitin, it was close to zero (0.8±3.6). For the mixture, the zeta potentials were -12.6±9.1 mV, -13.5±16.2 mV and -21.9±6.3 mV, respectively, for U:F ratios 5:1, 1:1 and 1:5. Thus, ubiquitin is more stable against aggregation in the presence of fullerols.

The binding constant of fullerol on ubiquitin was estimated at 4.6×10^4 M^{-1}. Using fluorescence quenching measurement, the binding site was located near the tyrosine residue. In the presence of fullerol nanoparticles, a small decrease in the α-helical and an increase in β-sheet contents were also observed using circular dichroism measurements. These measurements suggest association between fullerol and ubiquitin.

Effects on secondary structure of proteins

In protein folding, the favorable free energy contributions arise from interactions such as hydrophobic interactions, hydrogen bonds and van der Waals forces. These are countered by the entropic contribution which is unfavorable for the folded state. Since the energy balance between these two contributions is marginal for many proteins, the interaction with nanoparticles could induce structural changes. In many works, the effect of nanoparticles on the secondary structure has been studied. With fullerene derivatives, a decrease in α-helical content and an increase in β-sheet struc-

the electrostatic potential at the surface of the inner layer (or potential at radius of shear) with respect to a point in the bulk. Particles with large zeta potentials (absolute value) are, in general, more stable against aggregation.

tures was reported in many studies [111, 120, 126]. This trend is general and is seen for interaction of proteins with many other nanoparticles [127–129]. There have also been some studies in which fullerene derivatives showed no effect on the secondary structure of proteins [112]. The effects of nanoparticles have a curvature dependence: particles with low curvature (large size) cause more changes to the secondary structure compared to particles with high curvature (small size) [130]. Since preserving structure is vital for the functioning of proteins, functioning of protein is altered or even lost when nanoparticles induce structural changes.

4. Computational methods

Traditionally, theory and experiment are considered as the two branches of scientific research. Computational science is relatively a new branch of science even though it has been in existence for more than five decades. Today, it can be considered as the third branch of science, complementing the other two branches. The growth of the scope of computational science is caused by many developments [8]. First, the available computational power has exponentially increased as described by Moore's law[1]. Second, fast and efficient computational methods and algorithms have been developed or are being developed. For computational biophysicists, there is one more useful development: Due to the improvements of experimental techniques, structures of a large number of molecules are being resolved. Consequently, "a new enthusiasm is stirring in the molecular biophysics community" [8].

Physics-based computational methods use both quantum mechanical and classical approaches. Quantum mechanics-based computational methods are useful in studying phenomena involving chemical reactions, protein-ligand interactions and photo-induced reactions, to name a few. A classical approach suffices to describe many other phenomena. In these phenomena, quantum mechanical processes do not play an important role and hence classical approximation is possible. Self-assembly of lipid membranes, protein folding and interaction of proteins with small molecules and with other proteins (but not in all cases) are examples of such phenomena. The quantum to classical approximation is discussed in more detail in section 4.1.1. There are also hybrid methods (QM/MM) which treat the most interesting part of the system (*e.g.* enzyme reaction site) quantum mechanically and other parts classically

[1]Moore's law states that the number of transistors that can be fabricated on chips doubles approximately every two year.

[131, 132]. The advantage of such an approach is that it increases the speed of calculation without losing the accuracy.

In this Dissertation a computational approach is used to study the interaction of proteins with fullerene-based nanoparticles. The two computational methods used are molecular dynamics simulations and molecular docking (or simply, docking). Both of these methods treat atoms and molecules classically, but their methodologies are different and they serve different purposes.

4.1 Molecular Dynamics

In the molecular dynamics (MD) approach [1, 133], atoms are considered as point masses which follow classical equations of motion. The interactions between atoms are approximated by simple functional forms, and the positions and velocities of atoms are updated by numerical integration. The resulting trajectories, with the help of statistical physics, are used to calculate thermodynamic properties. MD is also useful to "see" molecular movements or conformational changes, which are often difficult or impossible to observe in experiments.

The MD methodology is used in a wide variety of research problems, and it is probably the most commonly used computational method in molecular biology. It is used to study all major classes of molecules in biology — lipids (membranes), proteins, carbohydrates and nucleic acids. The first MD simulation was the simulation of a hard sphere system performed in 1957 [134]. The first biomacromolecular simulation had to wait another two decades till 1977 when McCammon *et al.* carried out a 9.2 ps long simulation of bovine pancreatic trypsin inhibitor [135]. Today, millisecond-long macromolecular simulations are achievable [136].

The rationale behind the classical approximation and the equations of motion are described in the next two sections. The potential energy function is discussed in Section 4.1.3, and the selection of parameters in 4.1.4. Numerical algorithms used to integrate the equations of motion and other practical issues are discussed in section 4.1.5. Finally, in Section 4.1.6, some of the challenges faced by the MD community and the reasons why MD is still usable are discussed.

4.1.1 From quantum to classical

In principle, the behavior of any non-relativistic physical system can be described by solving the Schrödinger wave equation,

$$\hat{H}\psi_n = E_n\psi_n \tag{4.1}$$

where $\hat{H}$ is the Hamiltonian operator, ψ_n are the eigenstates, and E_n the corresponding eigenvalues. However, the wave equation can analytically be solved only for very simple systems such as the hydrogen atom. Solutions are thus based on many approximations and often calculated numerically [133].

The first approximation one makes is the so-called Born–Oppenheimer approximation, named after Max Born and Robert Oppenheimer who proposed it in 1927 [1]. It relies on the fact that the mass of electrons in an atom (or molecule) is much smaller than the mass of its nucleus, and hence the nucleus is fixed on the timescale of electron movement. The electron wave function depends only on the nuclear position (and not on momentum) and thus the electron wave function can be decoupled from the nuclear wave function. Solving the electron wave functions, one gets the so-called Born-Oppenheimer energy surface [8]. The nuclear wave function is then solved using the ground state electronic energy surface as a potential.

Although it is in principle possible to find numerical solutions to the Schrödinger equation for systems of interest, in practice for large systems such as proteins, even numerical solutions are cumbersome. Luckily, for studying many phenomena a complete quantum mechanical treatment is not necessary. In *molecular mechanics* (MM), the (average) atomic forces arising from inter-atomic interactions are represented by simple analytical potential energy functions. Both the functional forms of the potentials and the values of their parameters are required to completely describe an MM *force field*. The parameters in MM force fields are optimized against experimental data and/or quantum calculations [137].

4.1.2 Equations of motion

The basic concept of MD is very simple: the atoms behave classically and follow Hamilton's equations of motion

$$\frac{\partial H(\mathbf{p},\mathbf{q})}{\partial \mathbf{p}_i} = \dot{\mathbf{q}}_i \tag{4.2}$$

$$\frac{\partial H(\mathbf{p},\mathbf{q})}{\partial \mathbf{q}_i} = -\dot{\mathbf{p}}_i, \tag{4.3}$$

where the Hamiltonian of the system, $H(\mathbf{p},\mathbf{q})$, is a function of generalized coordinates $\mathbf{q}$ and generalized momenta $\mathbf{p}$,

$$H(\mathbf{p},\mathbf{q}) = \sum_i \frac{p_i^2}{2m_i} + V(\mathbf{q}). \tag{4.4}$$

The solutions are found successively in small steps of Δt using a finite difference approach. For solving, we should know the potential energy function $V(\mathbf{q})$ as a function of atomic coordinates and have initial positions and velocities of atoms.

4.1.3 Force field functional form

Force fields are descriptions of the potential energies of molecules as a function of their atomic coordinates [8]. The name comes from the fact that the forces acting on atoms depend on the gradient of the potential energy. The interactions between atoms are modeled by mathematical equations and appropriate constants for these equations are determined.

There are some conditions, though not rigorous, that force fields are expected to satisfy [8]. First, the force field terms have to be *additive*. That is, the effective energy of a molecule can be written as a sum of physical terms, such as those describing bonds, angles, electrostatic interactions etc. The terms should be *simple* to make the computation easy. But, at the same time, they have to represent the interactions satisfactorily. So there is usually a trade-off between accuracy and computational expense. The force field terms ideally have to be *transferable* between different

molecules, meaning the same parameters should be able to be used for closely related molecules. Force fields should also perform under a variety of environments and conditions. For instance, they should perform over a range of temperature and pressure and under different solvent conditions.

The interactions between atoms can be divided into non-bonded and bonded interactions. The non-bonded interactions include the dispersion interaction, repulsion at very close distances arising from Pauli exclusion principle, and electrostatic interactions. The first two terms are typically treated together using the 12-6 Lennard-Jones (LJ) interaction function [133],

$$V_{LJ} = \sum_{i,j} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \tag{4.5}$$

where i and j are atom indices and r_{ij} is the distance between them. σ is the collision diameter (distance at which the interaction energy is zero) and ε is the potential energy well depth. These two parameters are defined for each atom type and combination rules are used for calculating the interaction between a pair of atoms. For example, in some force fields, the ε value for a pair is obtained via geometric mean, $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$, and the σ value via the arithmetic mean $\sigma_{ij} = (\sigma_i + \sigma_j)/2$. In some other force fields, a geometric mean is used for both σ and ε. Mixing of parameters from different force fields is not recommended if the force fields follow different combining rules [138].

In Eq. (4.5) the sixth-power term represents the dispersion interaction which arises from fluctuating dipole-induced dipole interactions. It has a sixth-power inverse relationship with the distance between atoms [9]. The Pauli repulsion is related to wave function overlap, and can be effectively fitted against exponential functions of the distance between atoms [133]. However, for computational efficiency, it is usually replaced with a 12^{th} power term which can be easily calculated by squaring the dispersion term. This approximation works for most of systems, but for some systems, such as hydrocarbons, the 12^{th} power term is too steep [133]. The Buckingham potential uses an exponential term for the repulsive interaction and a sixth-power term

for the dispersion interaction [133],

$$V_{BH} = Ae^{-Br_{ij}} - \frac{C}{r_{ij}^6}, \tag{4.6}$$

where *A, B* and *C* are constants. At very small distances, the Buckingham potential is attractive and special care needs to be taken to avoid unrealistic atomic contacts [133].

The electrostatic interaction is represented by standard Coulomb's potential,

$$V_q = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}. \tag{4.7}$$

where q_i and q_j are the *partial charges* on atoms i and j, and ε_0 is the permittivity of free space. In molecules, the electrons involved in chemical bonds oscillate between bonded atoms. Thus, the effective charges of atoms (except fully ionized ones) are not integral multiples of e, the elementary charge unit (and hence the charges are called "partial charges"). To accurately represent the electrostatic nature of molecules, additional electrostatic interaction sites may need to be defined [133]. This is to ensure that higher order electrostatic moments are reproduced along with the lowest ones. The partial charge of an atom is not an experimentally measurable quantity. Hence the values are first obtained and then adjusted with other parameters (such as LJ parameters), such that the chosen thermodynamic quantities are reproduced.

In Eqns. (4.5)–(4.7), the summations are typically carried out over all inter-molecular atom pairs and those intra-molecular pairs which are separated by at least three bonds. For atoms separated by one or two bonds, the bonded terms are thought to represent the net interactions, so non-bonded terms are not necessary [138]. For those separated by three bonds, the normal LJ interaction may be too strong to distort the molecule [139], so it is scaled down by a factor in most force fields [137, 138]. The OPLS and AMBER force fields use a scaling factor of 0.5, but in the CHARMM force field the scaling factor is 1 (no scaling).

The bonded interaction function includes terms to model chemical bonds, bond angles and dihedral terms. These three terms describe energy costs for bond stretching,

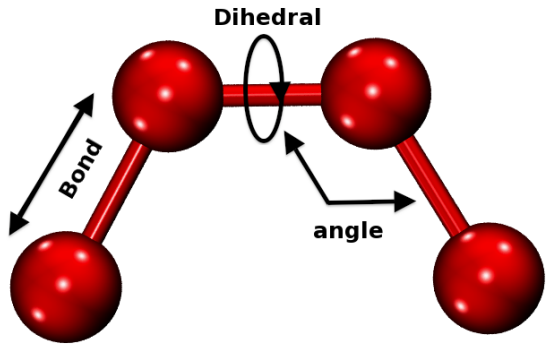


Figure 4.1. The bonded terms of molecular mechanics force fields.

angle stretching and rotation about bonds, respectively. The bonds and angles are usually represented by harmonic functions, dihedral terms using trigonometric functions with appropriate multiplicity (multiplicity is the number of minima in the dihedral angle potential as the bond is rotated through 360°). Thus the bonded potential function usually has four terms:

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{improper}. \tag{4.8}$$

Each term in the above equation can be expanded as follows:

$$V_{bonds} = \sum_{bonds} \frac{1}{2} k_{ij}^{b} \left(r_{ij} - r_{ij}^{0}\right)^{2}, \tag{4.9}$$

where r_{ij} is the distance between chemically bonded atoms i and j, r_{ij}^{0} is the reference bond length and k_{ij}^{b} is the force constant. The bond angle stretching term is

$$V_{angles} = \sum_{angles} \frac{1}{2} k_{ijk}^{\theta} \left(\theta_{ijk} - \theta_{ijk}^{0}\right)^{2}. \tag{4.10}$$

Here, θ_{ijk} is the bond angle defined by atoms i, j and k, θ_{ijk}^{0} is the reference bond angle and k_{ijk}^{θ} is the related force constant. The dihedral term is

$$V_{dihedral} = \sum_{dihedrals} k_{ijkl}^{\phi} \left(1 + \cos\left(n\phi_{ijkl} - \phi_{ijkl}^{0}\right)\right), \tag{4.11}$$

where ϕ_{ijkl} is the angle between *ijk* and *jkl* planes and ϕ^0_{ijkl} is its reference value. The OPLS force field uses a different form for the dihedral term,

$$V_{dihedral} = \sum_{dihedrals} \frac{1}{2}\left[C_1(1+\cos(\phi_{ijkl})) + C_2(1-\cos(2\phi_{ijkl}) + C_3(1+\cos(3\phi_{ijkl})\right]. \tag{4.12}$$

Additionally, improper dihedrals are used to force planarity of certain groups of atoms or to preserve chirality of molecules. They are called "improper dihedrals" because, unlike normal dihedrals, the torsion angle is defined by atoms that are not necessarily sequentially bonded.

$$V_{improper} = \sum_{improper} \frac{1}{2} k_{ijkl}\left(\xi_{ijkl} - \xi^0_{ijkl}\right)^2 \tag{4.13}$$

ξ_{ijkl} is the angle between *ijk* and *jkl* planes like ϕ_{ijkl}. However, unlike in proper dihedral term, here the atoms *i*, *j*, *k* and *l* are not necessarily connected via chemical bonds.

Other forms can also be used to represent the same potential. "Hard" degrees of freedom such as bond lengths do not deviate considerably from their minimum values. So the criterion for such potential functions is that they reproduce their minima very well. The bonds involving hydrogen atoms are often constrained since the vibrational energy is higher than the thermal energy[2], $h\nu \gg k_BT$ where k_BT is thermal energy, h is Planck's constant and ν is the vibrational frequency [1].

4.1.4 Force field parameters

In order to completely describe the potential energy surface, constants in Eqns. (4.5)–(4.8) have to be determined. At present, the main force fields used in biomolecular simulations are AMBER [140–143], CHARMM [144, 145], OPLS [146–148] and GROMOS [149]. The functional forms used in all these force fields are similar to

[2]Bonds involving hydrogen atoms vibrate with frequencies of about 3000–3500 cm^{-1}, so that $h\nu \approx 60 - -70 \times 10^{-21}$ J.

those described previously. The parameters for equations are determined in a self-consistent way to reproduce values from experiments or quantum chemistry calculations. Different force fields follow different philosophies and reproduce different thermodynamic quantities for parameterizing the constants. Data for small molecules is used for the parameterization because for large molecules, such as proteins, the properties of atoms may depend on its environment [150].

Partial charges are assigned to atoms to represent effective atomic charges. The charges are usually determined by quantum mechanical calculations on the molecule to be parameterized or on a closely related molecule [151]. These calculations are typically performed in vacuum and hence the charges often need to be adjusted further [138]. In an aqueous environment, the molecular dipoles are larger than in vacuum due to mutual polarization [138]. At present, the most commonly used force fields do not treat polarizability explicitly. Instead the values obtained from quantum chemistry calculations are overestimated to realistically model condensed phase systems [138].

The Lennard-Jones parameters are typically determined by reproducing condensed phase experimental quantities such as density, heat of vaporization and heat of sublimation [151]. The bond and angle values are determined from crystal structure values or vibrational spectra [8, 151]. Finally, the torsion terms are parameterized using quantum mechanical calculations on model compounds or vibrational spectra of molecules [8]. When a dihedral angle is changed, the energy profile will have a contribution from the non-bonded interactions between the end atoms (*i.e.* the 1-4 non-bonded interaction). Thus, in some force fields, the torsion and non-bonded terms are coupled and it is not recommended to transfer them separately from one force field to another one [138]. For proteins the ψ and ϕ dihedrals of the peptide backbone are of particular interest since they exist for all the residues and due to their significance in protein folding. These dihedrals are thus specially treated in some force fields [138]. The potential energy surface of these dihedrals are calculated for model systems such as alanine dipeptide using quantum chemistry methods [138]. The molecular mechanics parameters are then calculated by fitting dihedral functions to these quantum mechanical values.

Now let us discuss the two force fields used in this work, namely OPLS-AA [147]

and GROMOS 53A6 [149]. Of these two, OPLS is an all atom force field in which all atoms are explicitly modeled. GROMOS, on the other hand, is a united atom force field, where non-polar hydrogen atoms bonded to a carbon atom are merged to the carbon atom. The interaction parameters of the united atom is chosen such that it effectively represents the group of atoms.

OPLS-AA was parameterized by reproducing structural and thermodynamic quantities [147]. The bond and angle terms were mostly adopted from the AMBER force field (in AMBER these terms were determined using crystallographic data and normal mode frequency data for small molecular fragments [140]). The torsional terms were determined by fitting to energy values obtained from *ab initio* calculations. The non-bonded terms were determined by fitting thermodynamic and structural properties calculated from Monte Carlo simulations to experimental data. The thermodynamic properties used for fitting were heat of vaporization, density and free energy of hydration.

In GROMOS force field the covalent bonded terms were obtained from crystal structures of small molecules and the corresponding force constants from infrared spectroscopic data of small molecules in the gas phase [150]. The dihedral parameters were obtained by fitting functions to torsional energy profile obtained from quantum mechanical calculations. The first estimates of partial charges were obtained from quantum mechanical electron density values. Then these values were further optimized along with LJ parameters. For this step, experimental data for heats of vaporization, densities of liquids and free energies of solvation of small molecules in polar (water) and nonpolar (cyclohexane) solvents were used for fitting. Relative free energy of solvation between polar and nonpolar solutions is important for protein folding. For 20 amino acids the Gibbs free energies of solvation were calculated with the GROMOS 53A6 force field [150]. They were found to be very close to the experimental values, with average deviation smaller than thermal energy.

4.1.5 Numerical methods and practical issues

To get an atomistic trajectory, Hamilton's equations of motion (4.2) and (4.3) have to be solved numerically using an integrator. An integrator such as the leapfrog integra-

tor [152] calculates the position and velocities of atoms at a later time step from their values at the current (or earlier) time step, using the forces at the current time, *F(t)*:

$$\mathbf{v}\left(t+\frac{\Delta t}{2}\right)=\mathbf{v}\left(t-\frac{\Delta t}{2}\right)+\frac{\mathbf{F}(t)\Delta t}{m} \tag{4.14}$$

$$\mathbf{r}\left(t+\Delta t\right)=\mathbf{r}\left(t\right)+\mathbf{v}\left(t+\frac{\Delta t}{2}\right)\Delta t \tag{4.15}$$

An assumption made here is that forces do not change noticeably during the small time step Δt. Leapfrog algorithm is time-reversible and it conserves energy.

Since computational speed is an issue, it is desirable to use as large time step as possible. But using a large time step can lead to instabilities [133]. Bonds between atoms, which have high frequency and low amplitude, are the limiting factor for the integration time step. To avoid this, such bonds are replaced by constraints using algorithms such as SHAKE and LINCS. SHAKE [153] is an iterative algorithm which sets the correct bond lengths sequentially and it is numerically stable. But since bonds are coupled to one another, SHAKE has the drawback that a solution may not be reached within a fixed number of iterations. LINCS [154] solves constraint equations non-iteratively in two steps. It is more stable and up to 4 times faster than SHAKE [154].

Periodic boundary conditions

As will be discussed in Section 4.1.6, the maximum size of a system that can be simulated using MD is greatly limited. As a result, a large number of the simulated atoms are at the boundary of any system, unlike in reality. Since the behavior of atoms at the boundaries is different from those in the bulk, the finite size of the system can cause artifacts. To reduce these effects, MD simulations are generally performed with periodic boundary conditions (Fig. 4.2) [133]. Simulations are performed in a space-filling cell which is surrounded by its own translated images. For calculating short range interactions between atoms, a *minimum image* convention is followed: For calculating the short-range forces, only the distance between an atom and the nearest image of its interaction partner is considered.

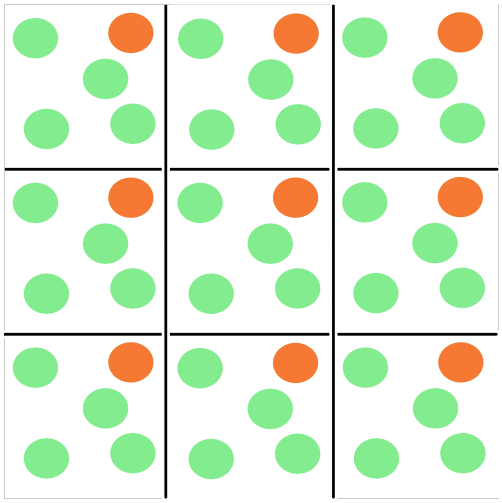


Figure 4.2. In a periodic boundary condition the simulated system is repeated in a periodic fashion to reduce finite size effects.

Calculation of electrostatic interactions can be problematic with periodic boundary conditions. Coulomb interaction is a long range interaction and hence its integral over all spaces diverges. Thus, special techniques have to be used to calculate electrostatic terms. The Ewald summation method was introduced by Paul Ewald for calculating the electrostatic energy of ionic crystals. It decomposes the interactions into short-range and long-range ones and calculates them in real space and the Fourier space, respectively [155, 156]. The advantage of the method is that the summation converges rapidly. Ewald summation method scales as N^2.

In the reaction field method [157], electrostatic interactions between atoms within a certain cutoff distance are calculated explicitly. Beyond the cutoff distance, the medium is treated as a continuum with a dielectric constant ε_m. This reduces the number of calculations that need to be performed.

In the particle mesh Ewald (PME) summation method [155], the charges are assigned to a grid using interpolation. The calculation is done in the Fourier space and forces are assigned back on to the atoms. The method scales as $N\log(N)$ [155] and is significantly faster than the ordinary Ewald summation.

Temperature and pressure coupling

Solution to Hamilton's equations (4.2) and (4.3) generate microcanonical ensembles (constant number of particles, volume and energy or constant NVE). Since experi-

mental measurements are typically made at constant temperature and constant pressure conditions, preferable ensembles for calculations in simulations are the isothermal (NVT) or isothermal-isobaric (NpT) ensembles [133].

In the Berendsen temperature coupling scheme [158], the system temperature is brought to the target temperature T_0 by changing the current temperature T at a rate

$$\frac{dT}{dt} = \frac{(T - T_0)}{\tau}. \tag{4.16}$$

The temperature change has an exponential dependance on the constant τ, whose value dictates how fast the system is brought back to T_0. Berendsen coupling, however, suppresses kinetic energy fluctuations and does not produce a canonical ensemble [159]. So it can not be used for computing quantities. But it rapidly brings the system to the target temperature and hence is often used for equilibrating the system before the production run. The stochastic velocity rescaling thermostat by Bussi *et al.* [159] generates the canonical ensemble and such thermostats are preferable for calculation of quantities. The velocities are rescaled to enforce the correct value of kinetic energy,

$$dK = (K_0 - K)\frac{dt}{\tau} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau}}, \tag{4.17}$$

where K and K_0 are the current and the target kinetic energies, respectively. N_f is the number of degrees of freedom and τ is the coupling constant. dW is a Wiener process [160] which ensures proper kinetic energy fluctuations. Nosé-Hoover thermostat [161, 162] also generates a canonical ensemble. However, in this approach the relaxation is oscillatory and hence it takes longer time scales to bring the system to equilibrium.

The principles behind pressure coupling schemes are similar to those of the temperature coupling schemes, but changes are applied to the pressure tensor $\mathbf{P}$ instead of temperature. In the Berendsen scheme [158], the pressure is maintained at the preferred value by changing the volume of the simulation box (and coordinates of atoms accordingly). But again, it does not produce the correct fluctuations and hence should only be used for equilibration purposes. For production runs, methods such

as the Parrinello-Rahman barostat [163, 164] are used.

Dealing with charges

Electrostatic monopole-monopole interactions are long-range and decreases only as the inverse first power of the distance between the monopoles. If periodic boundary conditions are applied to a simulation box with a non-zero net charge, its energy will diverge. Thus, with periodic boundary conditions, the total charge of the box should be made zero. The molecule that one wants to simulate may have a non-zero net charge. In this case, counterions are added such that the total charge of the simulated system is zero.

The ionic concentration inside a cell is different from the surrounding fluid. The typical concentration of ions inside cells is about 150 mM [165]. So for more realistic simulations, in addition to the counterions, more ions are added such that the system has an ion concentration of 100–150 mM. Na^+/K^+ and Cl^- are most commonly found cations and anions in cells [165] and hence are usually used in simulations.

4.1.6 Some of the challenges

There are many challenges facing the MD simulation community. One of the major challenges faced by MD, like many other computational methods, is its inadequate predictive power. MD should be able to accurately compute quantities that are experimentally measurable. This remains as a challenge especially in macromolecular simulation even after its four decades of existence [166]. This is partly because the potential functions representing inter-atomic interactions are sometimes not satisfactory. Another reason for the low predictive power of MD is that it is computationally very expensive. Consequently, the size of the system and length of simulation time are limited to a couple of millions of atoms[3] and/or a few μs. This finite size and length may result in insufficient sampling, which will affect the quality of "measurements" in MD simulations. Another disadvantage of MD is the limited transferability of force field parameters across different types of molecules [167], which limits its applicability.

[3]Compare it with 10^{14} which is the approximate number of atoms in an ordinary human cell. But vast majority of it is atoms in water molecules.

Despite these disadvantages, the use of MD has increased rapidly in the past. Almost all the major force fields used in biomolecular simulations are constantly developed, making the description of atomic interactions more and more satisfactory [166, 167]. Efficient simulation methods and algorithms are also developed. Replica exchange molecular dynamics [168], for example, enhances the sampling by simulating many copies of a system at different temperatures and then swapping the copies at regular intervals. This helps the samples to overcome barriers in the potential energy surface. Coarse graining is another method developed to speed up the simulations [169]. When a group of atoms is coarse-grained, it is represented by a single interaction bead. This reduces the number of interactions to be calculated in each MD step. It also makes the potential energy surface smoother allowing larger time steps. Thus a speed up factor of 1000 or more can be achieved by coarse graining [170]. Further, increase in computing power and massive parallel computing techniques have greatly extended the upper limit of simulation scales in the past two decades [166]. Thus upper limit of the length of trajectories generated in MD is also constantly increasing. As a result of all these developments, MD is becoming a tool for making accurate predictions [171].

4.2 Molecular Docking

Molecular docking is a method to study the association of two molecules. It predicts the binding modes and binding energies of one molecule with another one [172, 173]. Docking calculations are much faster than MD and less computationally demanding. But compared to MD, docking is less accurate and its scope is limited. It is not intended for calculating thermodynamic quantities or simulating dynamics of molecules. Docking is used in drug design to get binding "scores" for many potential drug molecules on a target protein [173]. The calculations are based either on the surface complementarity between the two molecules or on molecular mechanics based potential energy calculations [173].

Molecular docking is typically used to study the binding of small *ligand* molecules on larger *target* proteins, but it is also used to study protein-protein interactions [172, 174]. In docking simulations, a search algorithm generates a large number

of conformations for the complex and then energy functions are used to estimate the binding energy of each conformation. From the generated conformations and their binding energies, predictions of the complex structure are made. It should be noted that the goal of docking algorithms is to make a quick prediction, not to calculate the binding energies accurately. In the most simple case, both the ligand and target molecules are regarded as rigid bodies and only six degrees of freedom (three translational and three rotational degrees of freedom of the small molecule) are changed to generate conformations. But small ligands are, in most cases, highly flexible and hence are modeled as flexible molecules, treating the protein as a rigid molecule [172]. In more complex cases, both the ligand and target molecules are modeled as flexible. Often only the side chains of the binding site residues of the protein are modeled as flexible [172].

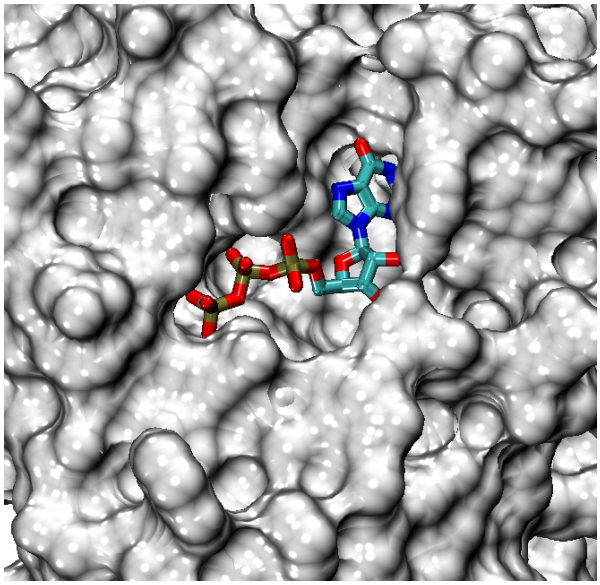


Figure 4.3. A small ligand molecule (GTP) is docked onto a target protein to find out the possible binding modes and estimate the binding energies. The figure was created using the molecular graphics software VMD [25].

4.2.1 Method

The search algorithms used in docking can be divided into three classes: systematic, stochastic and deterministic [173, 175]. In the first category, all the degrees of freedom are explored in a systematic way. Systematic algorithms may become cumbersome if the number of degrees of freedom is large. In deterministic methods (e.g.

MD simulation), the initial state determines the conformations that are subsequently generated. Deterministic methods have the disadvantage that if the complex reaches a local minimum on the energy surface, it may get trapped there and fail to continue the search [175]. In stochastic algorithms, the conformations are generated by making stochastic or random changes to the molecules. Since convergence is not guaranteed in stochastic methods, multiple copies of the simulations are performed [175].

In this work, a stochastic algorithm called Lamarckian genetic algorithm [176] was used for docking simulations. It is called a genetic algorithm because it resembles natural evolution. In this method, a population of potential solutions (in our case molecular conformations), which represent a population of phenotypes or individuals, is produced. Individuals are then *mutated* and a new generation is produced by *crossover* of individuals. In the mutation, small changes such as a change in a dihedral angle, are made to the conformation of individuals. To produce an individual of the next generation, parts of different individuals of the current generation are combined (crossover). A fitness function (described in Section 4.2.2) is used to evaluate the individuals and a certain number of the fittest individuals are allowed to evolve further in each generation. The search is ended when a certain number of generations are produced or when a satisfactory solution is reached.

4.2.2 Energy function

The docking simulations described in this Dissertation were performed using the AutoDock docking software version 4.2 [177]. The force field used in AutoDock for the prediction of free energy is semi-empirical and has been parameterized using a large library of protein-inhibitor complexes with known structures and inhibition constants. The binding energy can be written as [178]

$$\Delta G = \Delta U_{\text{lig}} + \Delta U_{\text{prot}} + \Delta U_{\text{prot-lig}} + \Delta U_{\text{conf}}. \tag{4.18}$$

Here the Δ's indicate that the terms are the difference in the energies between the bound and unbound states. The first two terms on the right-hand side represent the

intramolecular energy difference for the ligand and protein, respectively. If the protein or ligand is regarded as a rigid body, the corresponding term is zero. The third term represents the ligand-protein intermolecular energy difference and the last term corresponds to the entropy loss upon binding. $U_{\text{prot}-\text{lig}}$ is calculated using a pair-wise potential which has four terms representing van der Waals, hydrogen bond, electrostatic and solvation energy contributions [178]:

$$\begin{aligned} U_{\text{prot}-\text{lig}} &= W_{vdw}\sum_{ij}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+W_{hbond}\sum_{ij}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right) \\ &\quad +W_{elec}\sum_{ij}\frac{q_iq_j}{\varepsilon_{ij}r_{ij}}+W_{solv}\sum_{ij}(S_iV_j+S_jV_i)\,e^{\left(\frac{-r_{ij}^2}{2\sigma^2}\right)}. \end{aligned} \tag{4.19}$$

Here W_{vdw}, W_{hbond}, W_{elec} and W_{solv} are scaling factors which are parameterized by reproducing the known inhibition constants of a large library of protein-ligand complexes. The first term is the Lennard-Jones potential, the constants A and B were taken from the AMBER force field [179]. $E(t)$ in the second term is to make the hydrogen bond term directional. It is determined by finding the probability distribution of hydrogen bond angles from crystalline structures and then making a fit to the corresponding energy function [178, 180]. The parameters C and D are chosen for each type of hydrogen bond, such that the location and depth of the potential are reproduced. Atoms are assigned with Gasteiger partial charges [181] and the contribution is calculated using Coulomb's law in the third term. ε_{ij} is a distance-dependent dielectric term. The last term is the desolvation term which estimates the energy cost of desolvating from water. V is the volume of atoms surrounding a given atom and S is a desolvation term which estimates the cost of burying atoms during the complex formation from a fully hydrated state.

The entropy term in Eq. (4.18) has the most simple form and is directly proportional to the number of rotatable bonds, N_{tors},

$$\Delta U_{\text{conf}} = W_{\text{conf}}N_{\text{tors}}. \tag{4.20}$$

The parameter W_{conf} is calibrated along with the other W terms of Eq. (4.19).

4.2.3 Energy calculation

The molecular energy is rapidly calculated by using a grid-based method [182]. In this approach, prior to the docking calculations, a grid is embedded in the protein. Ligand atoms are placed at grid points and the interaction energy experienced at each grid point is calculated. During the actual docking calculations, the interaction energy of a ligand atom at any location is calculated from its values at grid points around that location by using trilinear interpolation.

In the grid based approach, as one would expect, the accuracy of calculation depends on the size of the gird — as grid size increases, accuracy decreases. Similarly, the speed of the calculation increases with an increase in the grid size. Instead of the grid based approach, a particle-based method like that used in MD is also possible and would be more robust. But compared to grid based methods, a particle-based method is significantly slower. Further, in docking simulations the goal is to make quick predictions and extremely accurate energy values are often not needed. Hence particle-based methods are usually not necessary.

4.3 Simulation details

4.3.1 Force fields and parameters

The OPLS-AA force field [146, 147] was used in all studies except in the tubulin-FUOH study. This choice of was based on the performance of the force fields and also some practical issues. In the earlier versions of the simulation package used (GROMACS, before version 4.5) GROMOS and OPLS-AA force fields were included, but not other prominent biomolecular force fields such as AMBER or CHARMM. Between the two available ones, OPLS-AA performs better. GROMOS, for example, favors β-sheet structures compared to other secondary structures [183]. OPLS-AA, on the other hand, has issues in reaching the folded state from an unfolded state, but once in a folded state, it preserves the protein structure very well [183]. Hence, OPLS-AA was used whenever it was possible. For tubulin-FUOH simulations, due to

the absence of appropriate parameters for GTP and GDP molecules in the OPLS-AA force field, the GROMOS 53A6 force field [149] was employed.

For proteins, the existing parameters in the respective force fields were used. The parameters for nanoparticles and GTP and GDP molecules were determined as described in the following. In conjunction with the GROMOS force field, the parameter choices made for FUOH molecules were as described in [109]. For non-bonded terms, bare carbon atom (type C) and default hydroxyl group parameters of 53A6 force field were used. Typical hydroxyl group partial charges of 53A6 force field were assigned to atoms belonging to the –C–OH groups. However, partial charges of the carbon atoms were reduced to 0.2188e from 0.266e to ensure charge-neutrality of the molecule. For all the other carbon atoms of the C_{60} core, the partial charges were set to zero. The parameters for GTP and GDP were obtained by combining the guanine parameters for the base and ATP molecule parameters for the rest of the molecules. Both of these parameter sets are available in GROMOS force field. The charge states of GTP and GDP were -4e and -3e, respectively. The partial charges on GTP and GDP atoms were also obtained from GROMOS force field. However, the terminal hydrogen atom was removed and the resulting charge was distributed on the oxygen atoms connected to the γ phosphate moiety [109].

In simulations employing the OPLS-AA force field, the bonded parameters for fullerene-based nanoparticles were adapted from the OPLS-AA force field. The Lennard-Jones parameters for carbon atoms belonging to the C_{60} core of the nanoparticle were adapted from Girifalco [184]. The partial charges of carbon atoms belonging to the C_{60} core were set to zero unless they were connected to any of the functional groups (hydroxyl group in FUOH, and malonic acid group in fullerene-trimalonic acid derivative). The partial charges for hydroxyl groups of FUOH were also obtained from the OPLS-AA force field. For charge neutrality of the molecules, slight modifications were made to the partial charge values of carbon atoms connected to the functional groups. For TMA, the partial charges for these carbon atoms were assigned by fitting point charges to a quantum mechanically determined electrostatic potential at the B3LYP/6-31G* level of the theory [185].

4.3.2 Practical implementation

The simulations were performed using the GROMACS simulation package [149]. Version 4.0.4 was employed for the simulation of *taq* DNA polymerase, and version 4.5 for tubulin simulations. Version 4.5.4 was used for all other simulations.

In all cases, the protein, protein-nanoparticle complex or the nanoparticle cluster was placed at the center of a simulation box whose edges were at least 0.9 nm from the solute. The box was then filled with TIP4P [186] (with OPLS-AA) or SPC [187] (with GROMOS) water molecules. Na^+ and Cl^- ions were added such that the system was charge neutral and had an ion concentration of 100 mM (In the case of tubulin, the ion concentration was 150 mM). However, in the case of the *taq* DNA polymerase simulations, ions were added only to make the system charge-neutral. Between each of the steps described above, the systems were energy minimized using the steepest descent algorithm implemented in GROMACS.

Each simulation was started with a 50 ps NVT equilibration simulation followed by a 100 ps NpT simulation. During these runs, the positions of heavy atoms were restrained. The temperature and pressure were maintained at their targeted values using the velocity rescaling algorithm by Bussi *et. al.* [159] and Berendsen coupling scheme [158], respectively. For production runs, the velocity rescaling algorithm [159] and Parrinello-Rahman barostat [163] were used for temperature and pressure coupling, respectively. In all cases, the solute and solvent molecules were coupled to separate thermostats.

In both equilibration and production runs, a simulation time step of 2 fs was used. The bonds were restrained using the LINCS algorithm [154]. The non-bonded interactions were calculated using a twin range cut-off scheme (cut-off distances of 0.9 nm and 1.4 nm). Electrostatic interactions were handled using the PME method [155, 156] with a cut-off of 0.9 nm. OPLS force field represents the folded states of proteins better when combined with PME compared to OPLS with straight cut-off [183]. Further, PME is a very fast and efficient method as described in 4.1.5.

4.3.3 Docking

The docking calculations were performed using the AutoDock automatic docking software version 4.2 [176]. The default AutoDock force field [182] was used for both protein and the nanoparticles. This force field is based on the AMBER force field as described in Section 4.2.

In all docking simulations, the conformation searching was carried out using a Lamarckian genetic algorithm [176]. The number of evaluations was fixed to 2.5×10^7. Since there was no previous knowledge of possible binding sites on any of the proteins, a blind docking was performed, allowing the algorithm to search all over the protein surface. The proteins were treated as rigid with no flexible backbone or side chains. The C_{60} core of the fullerene derivatives was also treated as rigid, but the bonds of the functional groups and those connecting the functional groups to C_{60} core carbons were modeled as flexible.

5. Results

This chapter discusses the results from the molecular docking and molecular dynamics simulations. In the first two sections, the cluster properties of fullerene-based nanoparticles and their effect on the secondary structures are described. In Section 5.3, the mechanism of inhibition of *taq* DNA polymerase activity by fullerene derivatives is studied. In Section 5.4, the inhibition of tubulin self-assembly by fullerol is studied. Finally, in Section 5.5, the effect of nanoparticles on ubiquitin is described. The simulation details including the force fields and parameters used are described in the previous chapter (Section 4.3).

5.1 Cluster properties

In this section the properties of small clusters of fullerene, fullerol and fullerene trimalonic acid are described. The systems simulated were (a) a single nanoparticle in water, (b) a cluster of 13 FUL molecules in an icosahedron configuration, and (c) 13 nanoparticles randomly placed in a box. The simulations were performed at 298 K. However, for studying the effect of temperature, the set c simulations were also carried out at a higher temperature (350 K). Six independent simulations were performed for each set, each simulation lasting 100 ns (except the first set of simulations which were only 5 ns long). In the following, first the results from simulations at 298 K are described which are then followed by the results from simulations at 350 K.

(a) A single nanoparticle in water

Description. A single nanoparticle was simulated in a box of water for the purpose of comparison. The temperature was kept at 298 K. The simulations were 5 ns long.

Results. The solvent accessible surface areas (SASAs) of nanoparticles were calculated. FUL is hydrophobic and the average surface area was 5.20 nm^2. FUOH and TMA have hydrophilic surface modifications. The total SASA of FUOH was 6.50 nm^2 of which 1.60 nm^2 and 4.90 nm^2 were hydrophobic and hydrophilic, respectively. TMA had a total area of 7.13 nm^2 of which 3.58 nm^2 was hydrophobic and 3.55 nm^2 hydrophilic. The FUOH molecule formed on average of about 34 hydrogen bonds with water molecules (see Table 5.2).

(b) 13 FUL molecules in an icosahedron configuration

Description. Fullerenes form fcc crystals with a lattice constant of 1.417 nm as mentioned in Chapter 3. However, in vacuum, small number of fullerenes form stable icosahedrons of 13 or 55 molecules [87]. To test stability of such clusters in water, a 13 FUL icosahedron structure was simulated in a box of water. Six simulations, each lasting 100 ns, were carried out. The temperature was set at 298 K. The results from these simulations are also important for the other systems described in the following sections. Icosahedron structure (Fig. 5.1) can be imagined of consisting of two pentagons of fullerenes (green and black) with a central fullerene (orange) and two caps (bright orange).

Results. The icosahedron structure remained very stable with no fullerenes dissociating in any of the simulations at any moment. The rms deviations of atomic coordinates were, however, rather large (about 0.5 nm). This is because the fullerenes, while maintaining the icosahedron configuration, were free to rotate about their centers of mass. The SASA per fullerene was 2.57 nm^2 in the icosahedron structure, which is half the value of an isolated FUL molecule. It should be noted that this number can be smaller for larger clusters with more fullerenes buried inside them. Hence this number itself is not very useful for describing the effect of clustering. However, it is still meaningful to compare this number for fullerene and its derivatives with different surface modifications. The radius of gyration of the icosahedron structure was 1.01 nm (standard deviation was smaller than 10^{-2} nm).

(c) Unstructured cluster of 13 molecules

Description. 13 FUL, FUOH or TMA molecules were placed randomly in a box of water. Six independent simulations were carried out at 298 K, each lasting 100 ns.

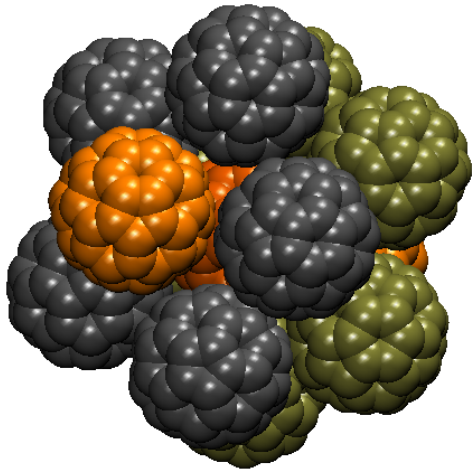


Figure 5.1. Fullerenes form icosahedral structures of 13 or 55 molecules. Here a 13 molecule icosahedral structure is shown. It can be imagine to consist of two pentagons (brown and dark green), one central molecule (orange) and two caps (bright orange).

The structures formed by different derivatives and their properties were studied.

Results. In all simulations, fullerene derivatives came close to each other and formed aggregates. However, the nature and stability of the aggregates formed were different for different derivatives. A typical structure formed by FUL molecules is shown in Fig. 5.2 and structures formed by FUOH and TMA at the end of six independent 100 ns simulations are shown in Fig. 5.3.

In all simulations, FUL molecules formed icosahedron-like structures with one or more FUL molecules at non-ideal locations. The reason for FULs to bind at non-ideal locations was that there are many different ways of arranging 13 FUL molecules to make an icosahedron structure. Different FUL molecules thus try to make it in different ways and this causes them to reach a deadlock. Since these non-ideal locations are local potential energy minima, FUL molecules do not easily rearrange themselves to make a more complete icosahedron. However, in one of the simulations, FUL molecules were seen rearranging themselves, forming a more complete icosahedron.

FUOH and TMA did not form icosahedron-like structures in any of the simulations. This was because they are not spherically symmetric like FUL molecules. FUOH and TMA molecules rather formed unstructured aggregates. TMA aggregates, unlike FUOH clusters, were very dynamic with their shapes and sizes changing over time.

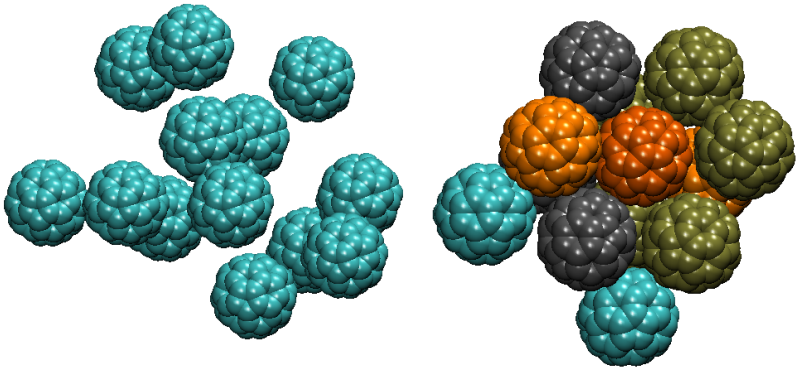


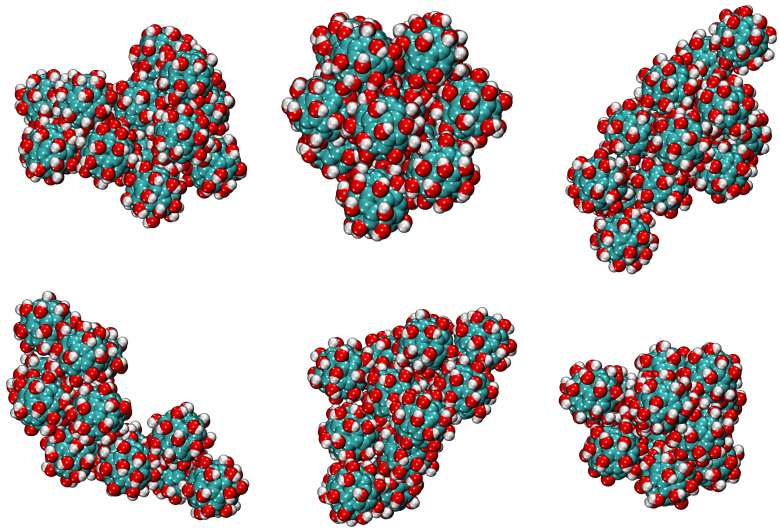
Figure 5.2. Typical structures formed by FUL molecules after 100 ns simulation (right) starting from a randomly arranged structure (left). The structure formed is icosahedron-like with (in this case) two cyan FUL molecules at "wrong" locations.

These two molecules lack rotational symmetry that FUL molecules have, and hence they were not seen rotating about their own centers of mass. This is because, in a bound state, a molecule requires to have rotational symmetry to be able to rotate freely about its center of mass.

The rms deviation of coordinates was calculated with respect to the structure at 100 ns. Choosing the last frame as reference structure was useful in distinguishing stable and unstable structures. For a stable cluster, the structure formed at the last time step and at a previous time would be very close to each other while for an unstable structure, they would be different. Hence, the rmsd values for the first case will be smaller compared to the second case. If the initial structure was used as the reference structure for rmsd calculations, both types of systems would produce large rmsd values making it impossible to distinguish stable and unstable structures.

For the FUL system the rmsd value was constant at about 0.5 nm (Fig. 5.4) because FUL molecules were free to rotate in the icosahedron-like structure. This value, as expected, was the same as the rmsd value of FUL molecules when the simulations were started by placing FUL molecules in an icosahedron structure (Section 5.1). TMA showed large rms fluctuations because TMA aggregates were very loosely formed and TMA molecules often disassociated from the cluster, occasionally forming more than one cluster. The stabilities of FUOH aggregates were somewhere between those of FUL and TMA. They had the smallest rmsd values because the structures were more stable than TMA, and they do not have the rotational degree of freedom like

(a) FUOH



(b) TMA

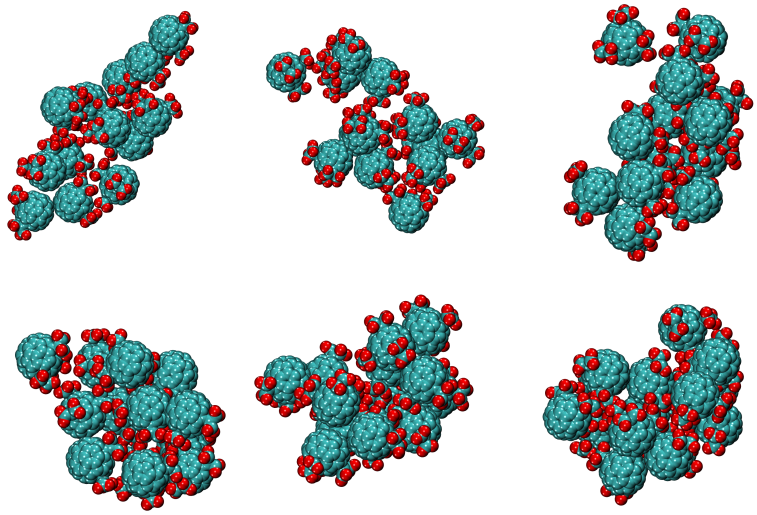


Figure 5.3. Structures formed by FUOH (above) and TMA (below) molecules at the end of six 100 ns simulations. Unlike FUL, these clusters were dynamic and lacked specific structures. The size and shape of the clusters changed with time.

SASA per molecule (nm^2)			
	Hydrophobic	**Hydrophilic**	**Total**
Single molecule			
FUL	5.20	0.00	5.20
FUOH	1.60	4.90	6.50
TMA	3.58	3.55	7.13
13 icosahedron (298 K)			
FUL	2.57	0.00	2.57
13 random (298 K)			
FUL	2.73	0.00	2.73
FUOH	0.97	2.87	3.84
TMA	2.79	3.00	5.79
13 random (350 K)			
FUL	2.72	0.00	2.72
FUOH	0.92	2.76	3.68
TMA	2.62	2.51	5.21

Table 5.1. Solvent accessible surface areas per molecule for different systems.

	FUOH-FUOH	**FUOH-Water**	**Total**
1 FUOH (298 K)		34.3	34.3
13 FUOH (298 K)	2.5	24.3	26.9
13 FUOH (350 K)	2.7	21.2	23.9

Table 5.2. Hydrogen bonds formed by FUOH molecules with themselves and with water molecules.

FUL molecules.

The average radii of gyration of the clusters were 1.09±0.08 nm, 1.28±0.20 nm and 1.62±0.20 nm for FUL, FUOH and TMA, respectively. TMA clusters had large effective radii of gyration, once again because of the loosely formed aggregates. TMA clusters also showed larger SASA compared to FUOH and FUL clusters. FUOH molecules formed a small number of hydrogen bonds with each other, but the number of hydrogen bonds formed with water molecules was reduced from 34.3 (for a single FUOH molecule in water) to 24.3 (see Table 5.2).

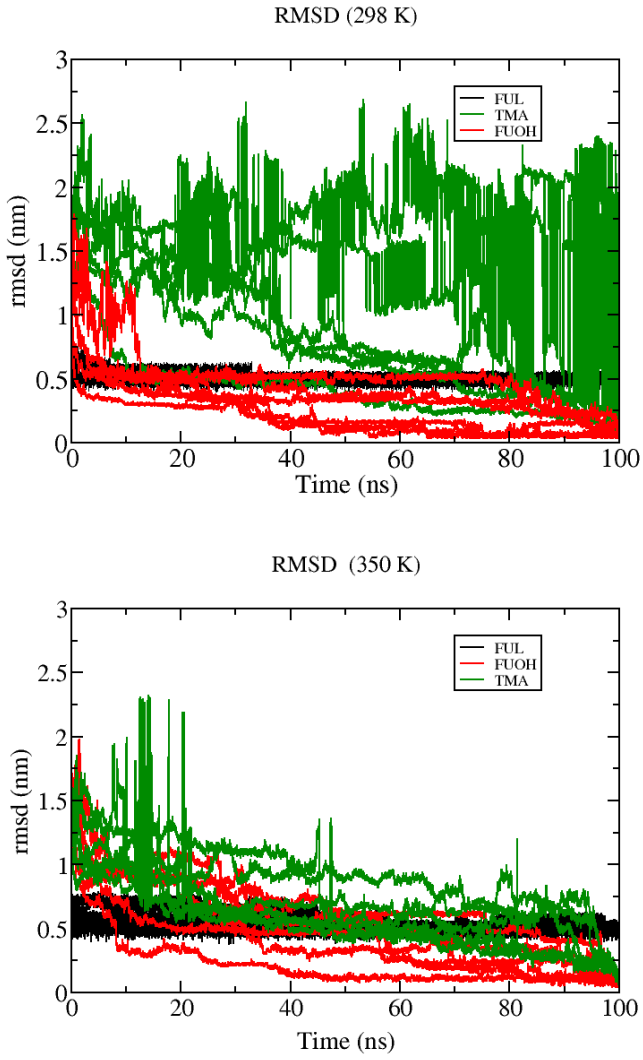


Figure 5.4. Root mean square deviation of atom coordinates from their final values at 298 K and 350 K. At 298 K the TMA structures were loosely formed which caused large fluctuations in rmsd. In comparison, at t 350 K, the clusters were more closely packed.

(d) Unstructured cluster of 13 molecules at 350 K

Description. The effect of temperature on cluster properties was studied by simulating the nanoparticles at a higher temperature. The simulated system was similar to the previous one, but the temperature was set at 350 K instead of 298 K.

Results. One would expect that the nanoparticles would aggregate less closely to each other at a higher temperature. However, on the contrary, the nanoparticles were more closely bound to each other at 350 K in comparison to simulations at 298 K. All calculated parameters showed this tendency for all the three nanoparticle types. Temperature had the greatest impact on TMA cluster behavior whereas the effect was very small for FUL. The radius of gyration values were reduced to 1.08±0.11 nm, 1.21±0.03 nm and 1.34±0.02 nm for FUL, FUOH and TMA at 350 K from 1.09±0.20 nm, 1.27±0.08 nm and 1.62±0.20 nm at 298 K. The rmsd values were also significantly smaller for TMA at 350 compared to 298 K values (Fig. 5.4), indicating the clusters were more stable at 350 K. Similarly, a small reduction in SASA was also observed (see Table 5.1).

Two complementing factors contribute to this temperature effect. The first one is that at higher temperature the dynamics are faster and the nanoparticles more effectively cross the barriers of the potential energy surface. Thus, the equilibrium structure is more easily reached. The second factor is based on the entropy cost of water molecules around the nanoparticle. When nanoparticles are placed in water, the water molecules around it can not rotate as freely as water molecules in the bulk. This causes them some entropy loss. When two or more nanoparticles come close to each other the total interfacial area between nanoparticles and water decreases and hence water molecules gain some of the lost entropic energy. This mechanism is the driving force behind self-assembly of many biomolecules such as lipids [188]. At higher temperature the entropy of water molecules is higher and so is the entropy loss due to nanoparticles. Thus, at higher temperature the entropic energy gain of water molecules up on nanoparticle aggregation is also higher (than at lower temperature), making the aggregation more favorable. A similar mechanism is responsible for the *cold denaturation* of proteins — for maintaining the protein structure, temperature has to be higher than a certain value [189, 190].

Peptide	Sequence
1ses	DREVQELKKRLQEVQ
1fkx	EEEKKELLERLYREY
5tim	PQQAQEAHALIRSWV
1trr	EEWLRFVDLLKNAYQ
3aaq	AAQAAAAQAAAAQAA

Table 5.3. The peptides simulated and their amino acid sequences.

5.2 Effect on secondary structure

The effect of fullerene nanoparticles on the α-helix propensities of small peptides was studied. To this end, a few peptides which were expected to form α-helical structures were simulated with and without FUOH and TMA. Each peptide was 15 residues long, and the amino acid sequences were obtained from the I-sites library of proteins [191]. These sequences have strong secondary structure correlations in the protein structures available in the Protein Data Bank. The amino acid sequences used are listed in Table 5.3. The starting helical structures were made by hand with molecular visualization program PyMOL [192], using its *Build* tool. The C- and N- termini of the peptides were modified by adding acetyl and amide groups, respectively. Five FUOH or TMA nanoparticles were placed around the peptide in complex simulations. The simulations were performed at 298 K and 1 bar. Eight independent simulations were carried out for each peptide/peptide-nanoparticle system. Each simulation was 25 ns long, providing a cumulative 200 ns long trajectory for each set.

Results. The number of residues involved in the α-helix formation was calculated for each simulation. The α-helix forming propensities of the peptides in the absence of nanoparticles varied. The peptide 3aaq had the lowest number of helical residues whereas 1fkx had the largest. The average number of helical residues for these peptides were 2.6±1.0 and 8.0±0.4, respectively. For the other three peptides, the helix forming propensities were in between these values (see Table 5.3).

In simulations with nanoparticles, they were seen binding to the peptides. In the presence of nanoparticles, no effect was observed for the peptides with high helix forming propensities. For these peptides, changes in the average number of helix forming residues were within the margins of error. However, for the two less structured pep-

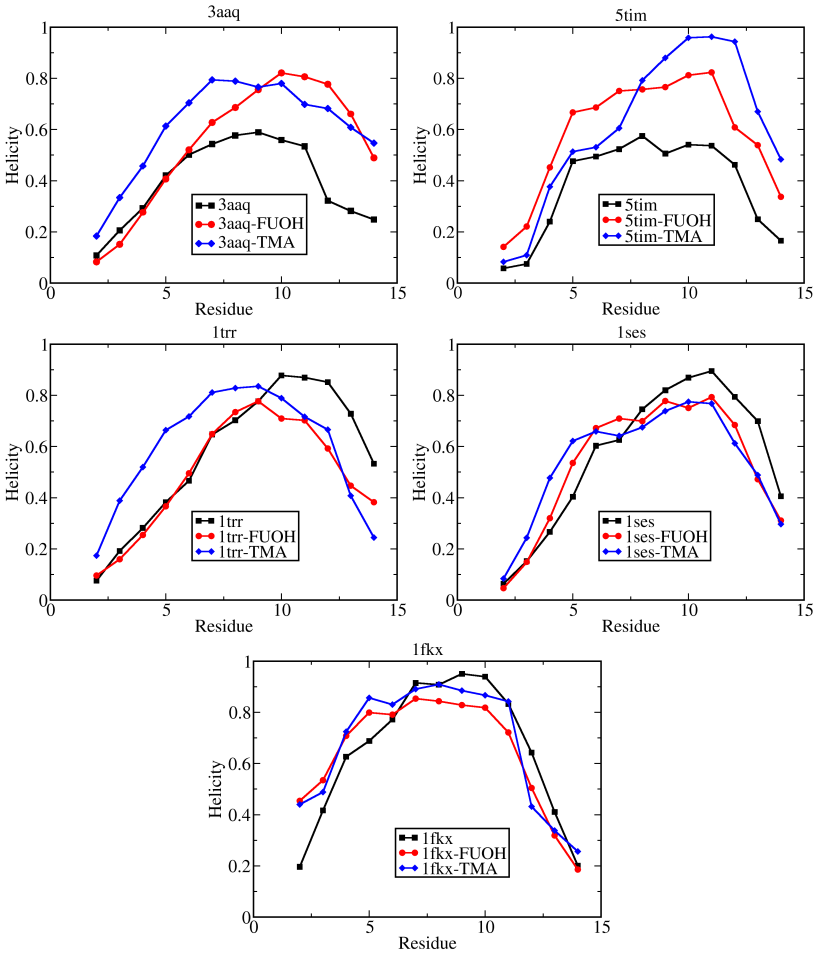


Figure 5.5. The average helicity of the peptide residues. The averages are taken over the last 10 ns of the simulations.

Peptide	**Number of residues**		
	No nanoparticle	**FUOH**	**TMA**
3aaq	2.6±1.0	5.0±1.0	6.2±1.2
5tim	4.2±1.1	6.8±0.6	7.7±0.5
1trr	5.7±0.6	5.7±0.7	7.1±0.9
1ses	6.6±0.4	6.1±0.8	5.9±1.1
1fkx	8.0±0.4	7.4±0.7	8.1±0.7

Table 5.4. The average number of residues forming helices for different peptides. The averages were calculated over the last 10 ns of the simulations.

tides (3aaq and 5tim) a significant increase in the helical propensities were observed upon nanoparticle binding. For 3aaq, the average number of helical residues were 5.0$\pm$1.0 and 6.2$\pm$1.2 with FUOH and TMA, respectively. For 5tim, the values were 6.8±0.6 and 7.7±0.5 in the presence of FUOH and TMA, respectively. The residue-wise helicities of the peptides are plotted in Fig. 5.5. For both these peptides, the prominent binding areas of nanoparticles were near the residues 5–12 which showed highest increase in helicity (Fig. 5.5). An analysis of rmsf values of protein residues showed a decrease in amino acid fluctuations upon nanoparticle binding. These findings indicate that the increase in helicity is induced by nanoparticle binding. However, further study is underway to assert this.

Discussion. The work presented here is very brief and requires further studies to understand the bigger picture. Future work directions include extending the present simulations up to 100 ns and studying the effects on β-sheet peptides or small proteins with β-sheet structures.

As described in Section 3.4, a general tendency for the helical content to decrease in the presence of fullerene-based nanoparticles has been reported in many experimental studies. In this study no such increase in the helical content was observed for any peptides. But on the the contrary, an increase in the helical content was observed for some peptides. Although these results are contradicting, increase in helical content induced by nanoparticles other than fullerene-based ones has been reported in the literature [193, 194].

The peptides simulated here are amphiphilic in nature. The stabilization of the helical structures by the nanoparticles was likely due to the hydrophobic interaction between the two. The hydrophobic areas of a protein are usually buried inside in its folded structure. So the binding of nanoparticles to these areas may look unlikely. However, it may still be important for proteins that follow diffusion collision model [195], that is, proteins in which secondary structures are formed first before they fold into a tertiary structure.

5.3 *Taq* polymerase-fullerene derivative interaction

In this section simulation studies of *taq* DNA polymerase (*taq* pol) with two fullerene derivatives are described. Motivation for this study was described in Section 3.4. Both docking and molecular dynamics (MD) approaches were used. Docking was used to find the most likely binding sites of fullerene derivatives on the protein. MD simulations of the most favorably docked structures and other systems were carried out to understand the mechanism by which fullerene derivatives affect the functioning of *taq* pol. Simulations were performed using only the Klen*taq* fragment [196] of *taq* pol. For brevity, in the following *'taq* pol' is used to refer to the Klen*taq* fragment of the protein.

5.3.1 Docking

Description. Two fullerene derivatives — fullerol and fullerene trimalonic acid — were docked onto both the open and closed structures of *taq* pol. The closed structure was in the DNA-bound form whereas no DNA molecule was included in the open structure. Both structures were obtained from the Protein Data Bank (PDB codes: 1TAQ [15] for the open and 3KTQ [12] for the closed structure). The 5' nuclease domain was cleaved from the 1TAQ structure by hand to get the Klen*taq* fragment. The nuclease domain was not included in the original 3KTQ structure. Water and other small molecules present in the PDB structure were removed. A hundred independent docking simulations were performed with 10 repeated trials in each simulation, making the total number of docked structures 1000 each for open and closed structures.

For each protein-fullerene derivative combination, the root mean square deviations (rmsds) between the docked structures were calculated. Structures with mutual rmsds below 0.8 nm were grouped into clusters. Each cluster represents a binding site of the fullerene derivative on the protein. Typical values used for the rmsd cut-off are in the range 0.05–0.2 nm for the docking of small molecules. But owing to their spherical symmetry of the C_{60} core, fullerene derivatives are able to bind the same site in different modes by rotating about their center of mass. The docked structures were clustered using different rmsd cut-off values, and the conformations within and across clusters were visualized using AutoDockTools [177]. Based on this trial and

error method, 0.8 nm was found to be the appropriate value which separates conformations corresponding to different binding sites without splitting the conformations corresponding to the same binding site into two or more clusters. From these binding sites, the most important sites were identified based on three criteria: (**a**) binding energy, (**b**) frequency of finding the site in repeated docking trials and (**c**) importance of the site in the functioning of *taq* pol.

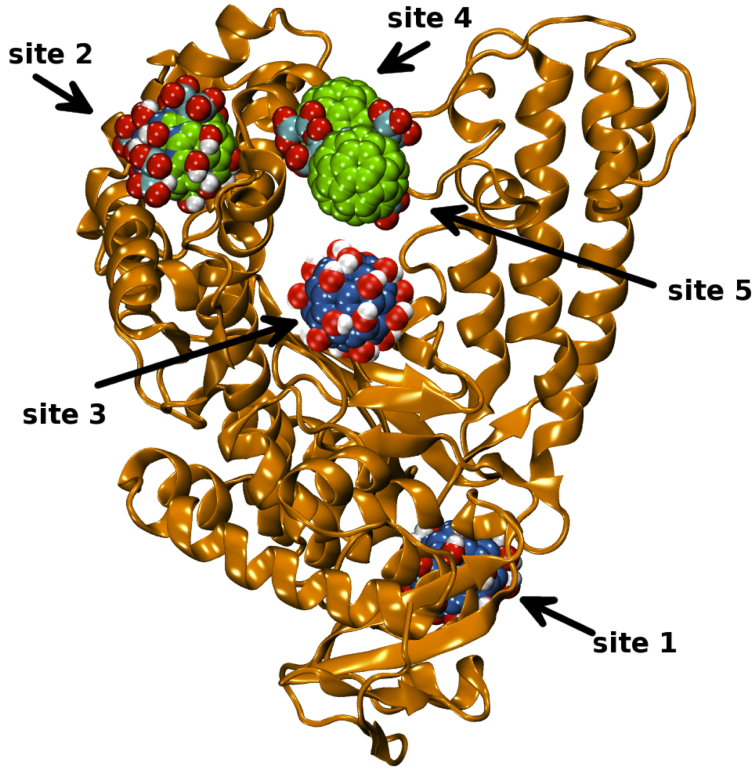


Figure 5.6. Binding sites of FUOH (blue C_{60} core) and TMA (green) on *taq* DNA polymerase in its open state, as predicted by the docking calculations.

Results and discussion. The important binding sites of the fullerene derivatives on *taq* pol are shown in Figs. 5.6 and 5.7. Their binding energies and contributions to total numbers of docked structures are shown in Table 5.5. In the open form of *taq* pol, FUOH had three important binding sites. Site **1** was located in the posterior side of the protein (with respect to the DNA binding region), site **2** was a pocket in the fingers subdomain and site **3** in the palm subdomain, near the active site of *taq* pol. In the closed state, FUOH was seen to bind sites **1** and **2**, and a new site formed by reorientation of the O-helix, site **6**. Site **3** was not a binding site in the closed form

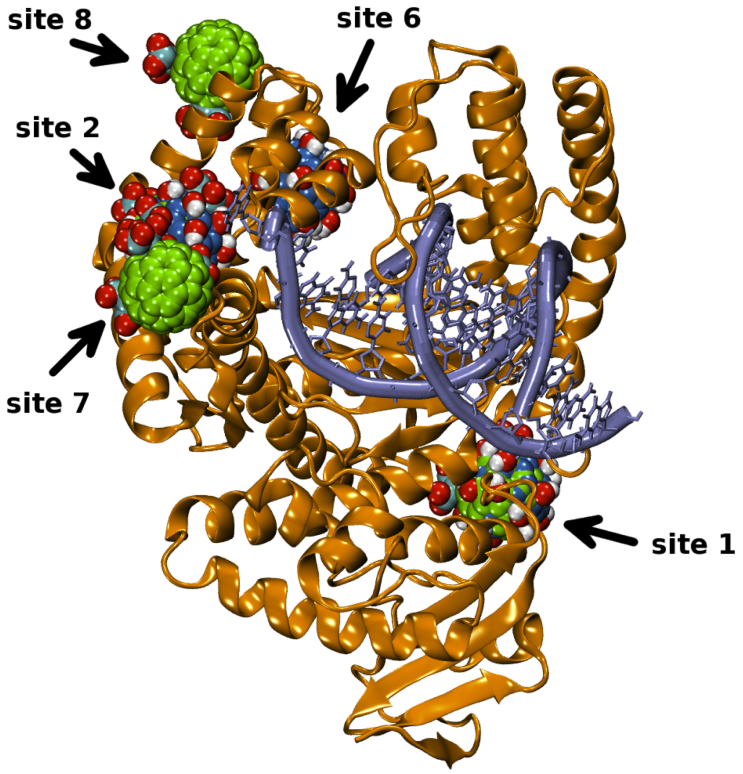


Figure 5.7. Binding sites of FUOH (blue C_{60} core) and TMA (green) on *taq* DNA polymerase in its closed state based on the docking calculations. The bound DNA molecule is shown in light blue.

because it is located in the same region where DNA binds to *taq* pol.

TMA bound to sites **1** and **2** in the open structure, but site **3** was not an important one. Two new sites, **4** and **5** were also favorable binding sites for TMA. When TMA was bound these two sites, it bridged the fingers and thumb subdomain (see Fig. 5.6). In the closed structure docking simulations, favorable sites for TMA were sites **1** and **2**, and two new sites, **7** and **8**. Sites **7** and **8** are located at the anterior and the posterior sides of the fingers subdomain, respectively.

The only site which was observed in both the open and closed structures and for both fullerene derivatives was site **2**. Site **1** was a favorable site for all cases except for TMA in the open form of the protein. Site **3** was in the DNA binding region. Hence, it was not a possible site when DNA is bound to the protein in the closed structure. Since the thumb subdomain is flexible, sites **4** and **5** are transient ones. Sites **6** and **8** are formed only when *taq* pol is in its closed form.

Site	E_b(kJ/mol)	Conf.(%)	Open/closed state	MD
FUOH, open state				
1	23.97	11.0	Both	•
2	19.67	55.8	Both	•
3	16.40	7.6	Both	•
FUOH, closed state				
1	21.09	10.5	Both	
2	23.68	58.0	Both	•
6	5.82	3.6	Closed	•
TMA, open state				
2	38.83	8.0	Both	•
4	48.12	38.3	Open	
5	35.94	37.2	Open	
TMA, closed state				
1	27.45	9.6	Both	
2	31.34	43.2	Both	•
7	27.07	29.3	Both	
8	26.99	7.1	Closed	

Table 5.5. Major binding sites of FUOH and TMA on the open and closed structures of the *taq* pol with their binding energies E_b and percentages out of all the docked configurations. The table also shows whether the binding site in question was found for the open, closed or both states of the protein, and whether the fullerene derivative-binding site pair was included in the MD simulation studies (•). The binding sites are shown in Figs. 5.6 and 5.7.

5.3.2 Molecular Dynamics

By using docking simulations the possible binding sites of the fullerene derivatives on *taq* DNA polymerase were found. To gain a more detailed understanding of the mechanism by which fullerene derivatives affect the structure and dynamics of *taq* pol, a molecular dynamics (MD) protocol was used. The simulation sets used were the following: In the APO set, the protein was simulated without any fullerene derivatives. In the SxF (or SxT) set, one FUOH (TMA) molecule was placed at site 'x' predicted by docking calculations. In R10F (R10T), *taq* pol was simulated with 10 FUOH (TMA) molecules at random locations around it.

The protein structures were obtained from the Protein Data Bank (PDB codes: 1TAQ for open structure and 3KTQ for closed structure) and were prepared as described in the previous section. The simulations were performed at 298 K and 1 bar. The polymerase chain reaction involves repeated heating and cooling of the reaction solution [197, 198]. To study the effects of temperature, a few simulations were performed at 333 K (simulation codes marked with an asterisk sign). During the polymerase reaction, *taq* pol switches between its closed and open configurations. Hence a few simulations of the closed structure (codes marked with a superscript C) were also

Code	System	#	Time (ns)
Open state (1TAQ), T = 298 K			
APO	Protein only	4	200, 3×50
S1F	1 FUOH at site **1**	3	3×50
S2F	1 FUOH at site **2**	3	3×50
S3F	1 FUOH at site **3**	3	3×50
S2T	1 TMA at site **2**	2	100, 50
R10F	10 FUOH randomly placed	6	2×100, 4×50
R10T	10 TMA randomly placed	6	2×100, 4×50
Open state (1TAQ), T = 333 K			
APO*	Protein only	4	2×100, 2×50
R10F*	10 FUOH randomly placed	3	100, 2×50
R10T*	10 TMA randomly placed	3	100, 2×50
Closed state (3KTQ), T = 298 K			
APO^c	Protein only	2	2×200
S2F^c	1 FUOH at site **2**	2	100, 50
S6F^c	1 FUOH at site **6**	2	100, 50
S2T^c	1 TMA at site **2**	2	100, 50

Table 5.6. Summary of the systems studied in the MD simulations. The table lists the codes assigned to each specific simulation set-ups, explanation of the set-up, number of independent simulations, and the simulation times of the independent runs.

performed. The typical lengths of simulations were 50-200 ns. Table 5.6 lists all the simulation sets along with their number of independent runs and lengths.

(a) Protein without any nanoparticles (APO)

Description. For comparison with the other simulations the protein was first simulated in its open state without any nanoparticles. Four independent simulations with a cumulative length of 250 ns were performed at 298 K.

Results. Visual examination revealed that the protein was in general very stable with no significant change in the structure. However, an exception to this was the tip of thumb subdomain which was flexible and mobile. This is not surprising because when DNA binds the protein this region undergoes a structural change. In this structural transition, the tip of the thumb makes a contact with the DNA and "holds" it. The root mean square fluctuations (rmsf) of residues, root mean square deviation (rmsd) of the C_α atoms from the crystal structure (Fig. 5.8), and DSSP analysis [199] of the secondary structure of the protein were conducted. The rmsf of the residues at tip of the thumb were higher than for the rest, as one would expect. No noticeable changes in the secondary structure was observed (see Fig. 5.11).

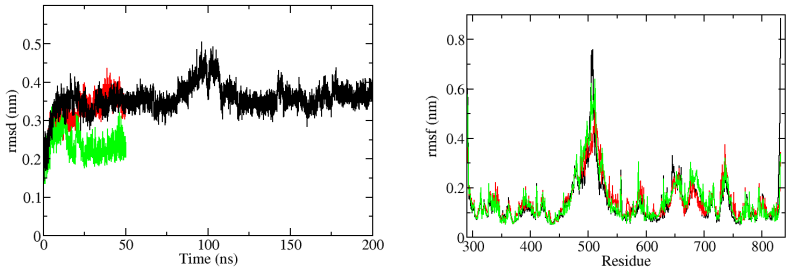


Figure 5.8. Root mean square deviation (rmsd) and root mean square fluctuation (rmsf) plots for *taq* pol without any nanoparticles. The rmsd is calculated using only the C_α atom coordinates. The crystal structure (1TAQ) was used for reference. The rmsf calculations were performed for each residue. Different colors represents plots for independent runs.

(b) One nanoparticle at a binding site (SxF/SxT)

Description. In this set of simulations, the protein was simulated with a single fullerene derivative placed at one of the binding sites predicted by docking. However, not all favorably docked structures were simulated. Three independent simulations were carried out for each nanoparticle-binding site combination.

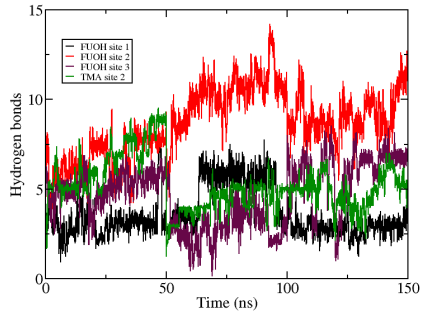


Figure 5.9. The number of hydrogen bonds that nanoparticles make at sites **1**, **2** and **3** as a function of time

Results. In all simulations, the nanoparticles remained at their respective binding sites during the entire simulation. This indicated that the binding sites predicted by docking simulations were indeed very favorable sites. However, in one exceptional case a fullerol was seen to move away from site **3** and bind to an adjacent site (called **3A** in the following) 1 nm away from it. The fullerol remained at this new site for the rest of the simulation. A new set of docking studies performed at site **3A** predicted a binding strength of 4.8 kJ/mol. This value is much smaller than the binding strength at site **3** (16.4 kJ/mol). When bound at the binding sites, fullerene derivatives made contact with many protein residues. Table **5.7** lists the residues that were within 0.5

nm of the fullerene derivatives at each binding sites studied.

The fullerene derivatives also formed a number of hydrogen bonds with the protein (see Fig. 5.9). On average, FUOH formed 3.6, 8.7 and 4.7 hydrogen bonds with the protein at sites **1**, **2** and **3**, respectively. At site **2** TMA formed an average of 5.7 hydrogen bonds. The average lifetime of the hydrogen bonds were 17.8±0.9 ps, 32.4±5.1 ps and 22.2±4.0 ps for FUOH at sites **1**, **2** and **3**, and 72.9±4.5 ps for TMA at site **2**. It is worth noting that in their small clusters FUOH and TMA formed smaller numbers of hydrogen bonds with each other (see Section 5.1). Table 5.8 lists the average number of hydrogen bonds the individual protein residues formed with the fullerene derivatives.

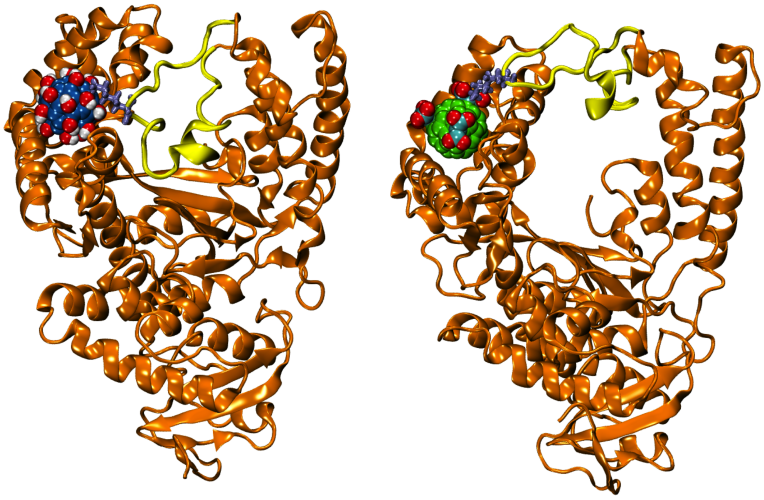


Figure 5.10. Tertiary structure change observed in S2F and S2T simulations. The tip of the thumb subdomain (yellow) moved towards and bound to the fingers subdomain.

In S2F and S2T simulations, a significant change to the tertiary structure was observed. The tip of the thumb subdomain moved towards the fullerene derivative which was bound at site **2** in the fingers subdomain (see Fig. 5.10). By doing so, the thumb residues Lys505 or Lys508 made contact with the fullerene derivative. For the rest of the simulation, the tip of the thumb subdomain remained bound to fingers domain.

Site-fullerene	Residues
1-FUOH	Met317, Ala360, Leu361, Arg362, Glu363, Gly364, Leu365, His443, Ala446, Thr447, Thr557, Arg559, His561, Glu773, Arg778
2-FUOH	Lys505, Lys508, Gly672, Gln 690, Ile693, Glu694, Phe697, Arg704, Ile707, Glu708, Leu711, Arg715, Arg741, Glu742, Glu745, Arg746, Phe749
3/3A-FUOH	Thr569, Ala570, Thr571, Arg573, Cysh575, **Asn580**, Gln582, **Asn583**, Ile584, Pro585, Val586, Phe667, Leu670, Tyr671, Arg746, Met747, **Asn750**, **Gln754**, His784
2-TMA	Lys505, Gly672, Met673, Ser674, Gln690, Ile693, Glu694, Phe697, Arg704, Ile707, Glu708, Leu711, Arg715, Lys738, Arg741, Glu742, Glu745, Arg746, Phe749

Table 5.7. *Taq* DNA polymerase residues in contact with the fullerene derivatives at binding sites **1**, **2**, and **3**. The residues forming hydrogen bonds with DNA at the active site are highlighted in boldface.

Site 1 - FUOH		**Site 2 - FUOH**		**Site 2 - TMA**		**Site 3/3A - FUOH**			
Residue	#HB	Residue	#HB	Residue	#HB	Residue	#HB	Residue	#HB
Leu 361	0.994	Lys 505	0.017	Lys 505	0.165	Thr 385	0.121	Gln 582	1.160
Gly 364*	0.037	Lys 508	0.140	Ser 674	0.025	Asn 565	0.106	Asn 583	0.213
Leu 365	0.023	Gly 672*	0.278	Ala 675*	0.159	Ala 568*	0.206	Arg 728	0.024
His 443	0.198	Gln 690	0.562	Gln 690	0.152	Thr 569	0.197	Asn 750	0.449
Ala 446*	0.510	Glu 694	0.168	Arg 704	1.451	Ala 570*	0.285	Gln 754	0.319
Thr 447	0.931	Arg 704	0.695	Arg 715	0.826	Thr 571	0.767	His 784	0.200
Glu 773	0.496	Glu 708	1.458	Lys 738	0.791	Arg 573	0.266		
Arg 778	0.409	Arg 715	0.147	Arg 741	0.674	Cys 576	0.154		
		Arg 741	0.079	Arg 746	0.964	Asp 578	0.089		
		Glu 742	1.235			Asn 580	0.090		
		Glu 745	2.016						
		Arg 746	1.095						

Table 5.8. Average numbers of hydrogen bonds (#HB) formed by FUOH and TMA with the residues at the binding sites **1**, **2** and **3**. A hydrogen bond with the backbone is designated by an asterisk after the name of the residue.

(c) Many nanoparticles around protein (R10F/R10T)

Description. In the R10F and R10T simulations, ten FUOH or TMA molecules were placed at random positions around the protein. Six independent runs were simulated with a cumulative time of 400 ns for R10F and R10T each. The initial positions of the fullerene derivatives were changed in each simulation for better sampling. The simulations were performed at 298 K on the open structure.

Results. The fullerene derivatives immediately moved towards the protein surface and remained bound to the surface for the rest of the simulation. There were no particularly favorable sites for fullerene derivatives. However, TMA showed a tendency to bind to positively charged residues because of its large negative charge (-6*e*). FUOH, on the other hand, had a tendency to bind to polar or charged residues. Both fullerene

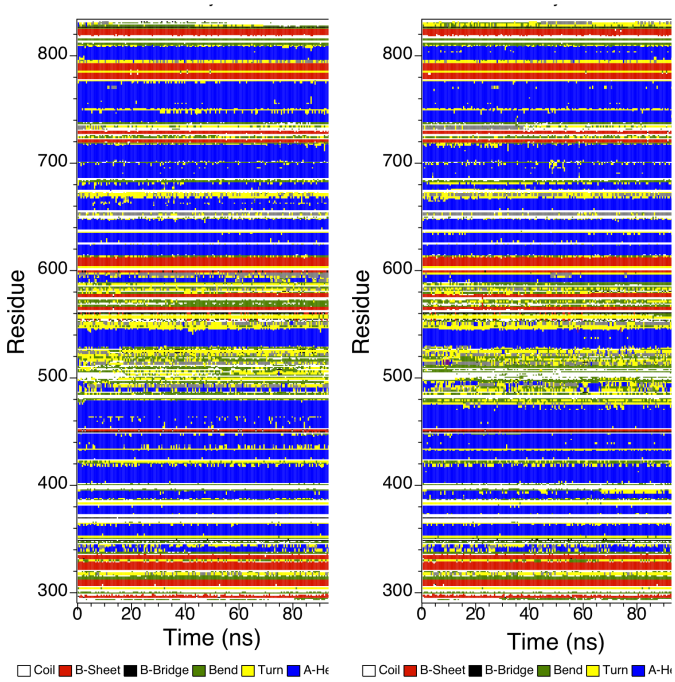


Figure 5.11. DSSP analysis showed no significant change to the secondary structure of *taq* pol in the presence of nanoparticles. In this figure, the secondary structures formed by *taq* pol residues as a function of time for a simulation without any nanoparticles (left) and one simulation with 10 FUOH nanoparticles (right) are shown.

derivatives formed hydrogen bonds with the protein residues as expect. Each FUOH molecule formed on average of 3.5 hydrogen bonds while TMA made an average of 2.6 bonds. These numbers are smaller than the numbers of hydrogen bonds formed at the binding sites predicted by docking. This also indicates that the sites predicted by docking simulations were very favorable binding sites for the fullerene derivatives. The hydrogen bond lifetimes were 24.2±3.2 ps and 49.0±9.9 ps for FUOH and TMA, respectively.

Further, in some simulations fullerene derivatives were bound the protein at the binding sites predicted by docking simulations. However, they were not seen to bind to all the bindings sites predicted by docking calculations. This is likely due to the sampling limitations of MD. The protein is rather large and the longest simulations were limited to 100 ns due to computational cost. The sampling was however enhanced to a certain degree by using 10 fullerene derivatives in a single simulation. It is worth mentioning at this point that the binding sites are not always well defined in MD as in the static docking calculations. This is because in MD the protein is flexible and consequently changes to binding sites may occur. This is especially true for the tip

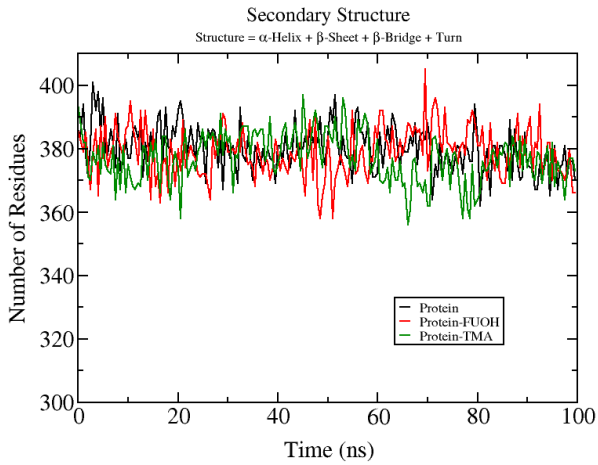


Figure 5.12. The number of residues of the protein forming secondary structures in the presence and in the absence of fullerene derivatives.

of the thumb subdomain.

The secondary structure of the protein was analyzed using DSSP[1] [199]. Fullerene derivatives did not seem to make any significant change to the secondary structure. In Fig. 5.11, the secondary structures formed by each residue is plotted as a function of time for both protein-alone and protein with 10 FUOH simulations. In Fig. 5.12, the number of residues that form secondary structures is plotted as a function of time for the protein with and without FUOH and TMA.

Binding of the fullerene derivatives at certain locations caused significant changes to the protein tertiary structure. In some simulations, the fullerene derivative was seen to bind to two α-helices (H-helix, residues 453–477, and I-helix, residues 527–552, see Fig. 2.6) of the thumb subdomain. This caused these helices to bend, as depicted in Fig. 5.13. This tertiary structure change was not observed in any of the simulations where the nanoparticles did not bind to the H and I helices.

Another kind of tertiary structure change was seen in the tip of the thumb subdomain. Similar to observations in the S2F and S2T simulations, in some simulations this region moved towards the fingers subdomain. In such occasions, the fullerene derivatives were seen to bridge the fingers and thumb subdomains by binding to both of them.

[1] DSSP is an algorithm to assign secondary structures to a protein structure. The algorithm is based on the intra-backbone hydrogen bonding patterns and energies.

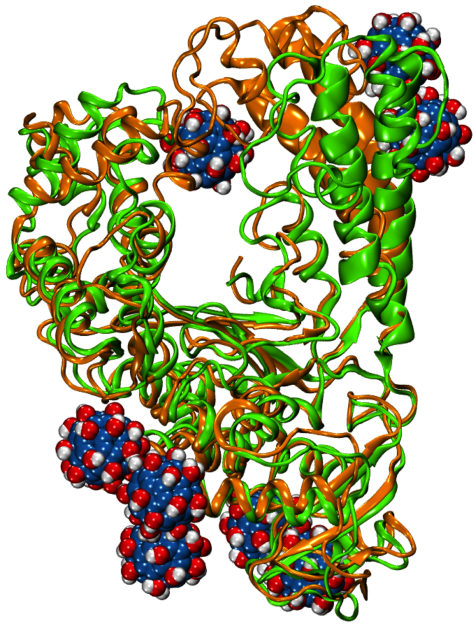


Figure 5.13. In simulations in which FUOH bound the H and I helices of the protein, a bending of these helices. In other simulations in which no FUOH molecules bound to these helices, no such structural changes were observed. The crystal structure is shown in green and simulated structure in orange.

We defined two groups of residues in the fingers and thumb subdomains and calculated the distance between their centers of mass for the APO, R10F and R10T simulations. The chosen residues were 496-510 in the thumb and 673-690 and 732-742 in the fingers subdomains. The average value of this distance in the last 10 ns of simulations was 2.6±0.3 nm for APO, 2.2±0.4 nm for R10F and 1.5±0.2 nm for R10T simulations. The error estimates were made by finding the standard deviation of values between independent runs. In one R10F simulation none of the FUOH molecules bound to the tip of the thumb. If this simulation was excluded from the analysis, the average distance was 2.0±0.2 nm.

(d) Many nanoparticles, at T=333K (R10F/R10T*)*

Description. 10 fullerene derivatives of same type (FUOH or TMA) were placed around the protein, and three independent runs of FUOH and TMA simulations were carried out. The cumulative time of the simulations was 200 ns for FUOH and TMA each. The simulations were carried out at a temperature of 333 K. The motivation for using a higher temperature simulations was the fact that the polymerase chain reaction is usually set up at a temperature significantly higher than the room temperature.

Results. The protein showed no significant change in structure or stability. The rmsd values of the C_α atom coordinates were 0.40±0.04 nm, 0.52±0.12 nm and 0.37±0.01 nm for protein alone, protein-FUOH and protein-TMA simulations, respectively. The average number of hydrogen bonds FUOH and TMA formed with the protein at 333 K were 3.7 and 2.7 respectively. These are smaller than the average values at 298 K. The lifetimes of hydrogen bonds were 20.3±0.2 ps and 40.5±1.8 ps for FUOH and TMA, respectively. The distance between the groups of residues described in the previous section were 2.3±0.3 nm, 1.9±0.2 nm and 1.4±0.1 nm for protein alone, protein-FUOH and protein-TMA simulations, respectively.

(e) Closed structure (APO^c/SxF^c/SxT^c)

Description. The closed structure of the protein was simulated without the bound DNA molecule. The motivation for conducting this simulation was to check whether the closed-to-open transition would be affected by fullerene derivative binding. The systems simulated were *taq* pol without any fullerene derivatives, *taq* pol with FUOH or TMA at site **2** and *taq* pol with FUOH at site **6**. The temperature was set at 298 K, and two independent runs of simulations were performed in each case.

Results. In the simulations without any fullerene derivatives (APO^c), we expected that the protein would go from the closed to open state. However, in neither one of the two 200 ns simulations such a transition was observed. The reason for this could be that the closed structure is a local minimum even in the absence of DNA. In that case more independent of simulations would be required to see the closed-to-open transition and to compare it with the protein-fullerene derivative complex cases.

In one of the simulations with FUOH at site **2** ($S2F^c$), the protein made a transition from the closed to open state. This indicates that FUOH binding does not prevent the closed-to-open transition. However, no such transition was observed in any of the other protein-fullerene derivative complex simulations.

5.3.3 Possible inhibition mechanisms

In the following possible inhibition mechanisms of *taq* pol by FUOH and TMA are considered one by one. The reasons for accepting or rejecting them are discussed.

Finally some experimental methods for testing the proposed inhibition mechanisms are discussed.

Direct binding at the active site. The active site of the protein consists of three acidic residues. TMA is highly negatively charged and so it is not expected to bind to the negatively charged active site. However, FUOH is polar and it can, in principle, bind to the active site making hydrogen bonds with the acidic side chains. However, in neither docking nor in MD, the fullerene derivatives were seen to bind to the active site of the protein. The reason for this is that both FUOH and TMA are too large to fit into the active site which is located in a cleft in the palm subdomain. So it is unlikely that the fullerene derivatives inhibit *taq* pol by binding directly at the active site.

Binding at sites 3 and 3A. FUOH was seen to bind to sites **3** and **3A** of the protein. These two sites are located in a region in the palm subdomain where DNA binds to the protein. FUOH made contact with many protein residues when bound to these sites. At site **3** FUOH made hydrogen bonds with residues Ala570, Thr571, Gln582, Asn583, Asn750 and Gln754. Similarly, at site **3A**, FUOH made hydrogen bonds with Thr385, Ala568, Cys576, Cys577, Asp578 and Asn580. When a DNA molecule binds the protein, it makes contact with many residues including Ala568, Cys577 (Ser577), Asn580, Asn583, Asn750, and Gln754 [200]. Thus, FUOH binding at sites **3/3A** would directly affect the DNA binding and hence the activity of the protein.

However, TMA did not bind to sites **3** or **3A** in any of the simulations and cannot inhibit the protein with this mechanism. Further, for FUOH such inhibition would be competitive. Since FUOH and DNA cannot bind to the protein simultaneously, they would compete for the binding site. Then, if the concentration of DNA (or FUOH) is increased, the inhibition effects should decrease (increase). But this is not the case. In PCR experiments when the concentration of DNA was increased, no significant reversal of inhibition was observed [121]. Thus we conclude that although binding of FUOH at sites **3/3A** can contribute to the inhibition, it cannot be the only (or major) inhibition mechanism.

Tertiary structure changes due to binding. In some of the simulations with many FUOH/TMA particles around the protein (R10F/R10T simulations), the protein tertiary structure changed significantly. The thumb subdomain has two long α-helices,

the H-helix (residues 453–477) and the I-helix (residues 527–552), see Fig. 2.6. In some simulations, binding of FUOH at these two helices caused them to bend as shown in Fig. 5.13. Such bending was not observed in any of the simulations without FUOH or in simulations where the fullerene derivatives did not bind these helices. Interestingly, the I-helix is involved in DNA binding and makes a number of contacts with a *taq* pol-bound DNA.

There was also a second kind of tertiary structure change observed in simulations with fullerene derivatives. In a few simulations with FUOH or TMA bound at site **2**, the tip of the thumb subdomain moved towards the fingers domain. Two lysine residues of the thumb strongly bound to the fullerene derivatives preventing the thumb from going back to its original position. Similar thumb-to-fingers movements were also observed in simulations with many fullerene derivatives placed around the protein. In these cases, the fullerene derivatives acted like a bridge by binding both to the tip of the thumb and fingers subdomains.

Changes in the tertiary structure described above would affect the DNA binding in a non-competitive fashion. Although at higher DNA concentrations the availability of DNA would increase, the *taq* pol would not be able to bind them. This means that if the concentration of DNA was increased, the inhibition would not be reversed. On the other hand, if the concentration of *taq* pol was increased keeping the same concentration of the fullerene derivatives, there would be more unaffected protein molecules. Thus an increase in enzyme activity could be expected with an increase in the concentration of the protein. This is indeed exactly what was observed in experiments.

Point mutation studies to check the predictions. In point mutation studies one or more specific amino acids of a protein are replaced with different amino acids. This allows one to investigate the effect of that particular amino acid in the functioning or disfunctioning of the protein. In our case, two of the inhibition mechanism described above are based on the binding of fullerene derivatives to two thumb helices, tip of the thumb and at site **2**. The binding to the two helices was not very specific to any residues. Further, the tip of the thumb and one of the helices (I-helix) have very conserved residues which are involved in DNA binding. Mutation studies on these residues are thus very difficult.

Many residues at site **2**, on the other hand, are not conserved and hence would be suitable for mutation studies. It is also important to note that FUOH and TMA bind to different residues at site **2** (see Table 5.7). Residues that dominantly bound to FUOH were Glu708, Glu742, Glu745 and Arg746. Among these only Arg746 was seen to bind TMA. Arg746 is also absolutely conserved in the polymerase family. Glu708 and Glu742 are not conserved, whereas Glu745 is somewhat conserved. Thus the residues Glu708 and Glu742 would be optimal candidates for mutation studies.

5.4 Tubulin-fullerene derivative interaction

5.4.1 Docking

Description. FUOH molecule was docked onto a tubulin dimer to find the most probable binding sites. The crystal structure of the tubulin dimer with bound GDP and GTP was obtained from Protein Data Bank (PDB ID: 1JFF [26]). The missing residues of this structure were added using the program MODELLER 9v7 [201, 202]. Docking simulations were performed using the AutoDock 4.2 software [177] with its default parameters [182]. The C_{60} core of FUOH was treated as rigid and the C-O bonds connecting the hydroxyl groups to the C_{60} core were modeled as flexible. 50 docking simulations were performed with 10 trials in each simulation. The number of evaluations was set at 2.5×10^7.

In reality the proteins are flexible molecules and can adjust the relative positions of atoms and residues. Thus it is desirable to perform docking on conformations generated *e.g.* by MD simulations. However, since tubulin is a dimer molecule, in MD simulations the two monomers can change their structures independently. Even small such changes would make it difficult to simultaneously fit both monomer conformations (generated by MD) to their crystal structures. We found that this causes a practical problem in docking while defining the binding sites (which is based on rms deviations of the docked FUOH structures). As a trade-off between this practical difficulty and desirability to use multiple protein conformations, we used two structures

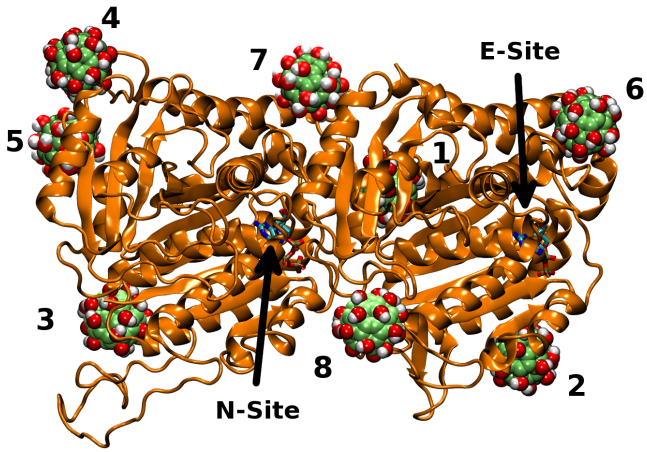


Figure 5.14. Eight binding sites of FUOH molecule on a tubulin dimer identified by docking simulations.

in docking simulations–the crystal structure and a conformation tubulin formed after 10 ns of MD simulations.

Results. As in the case of *taq* pol, 1000 docked conformations (500 for each protein conformation) were grouped into different clusters based on the mutual rmsds. Eight binding sites were identified as the important ones based on the binding energy, population of the clusters and the importance of the sites. The binding sites identified are shown in Fig. 5.14, and the binding energies and populations of these clusters are summarized in Table 5.9.

It is worth mentioning at this point that in a microtubule, the neighboring tubulins are connected either laterally (on the left/right sides in Fig. 2.8) or longitudinally (on top/bottom sides in the same plane in Fig. 2.8). Thus, most of the predicted binding sites are located at either interdimer or intradimer interfaces: sites **1**, **7** and **8** are at the intradimer interface, whereas all other sites are located at the interdimer interface.

5.4.2 Molecular dynamics

The binding of FUOH could affect the secondary, tertiary or quaternary structure of the protein. To examine this, two sets of MD simulations (tubulin dimers with or without FUOH molecules) were carried out.

Site	Lowest binding energy kJ/mol	Population	Residues within 0.7 nm of FUOH
1	27.2	85	Asn102, Glu411, His406, *Arg158*, *Asp163*, *Asn197*, *Thr198*, *Asp199*, *Val257*, *Phe262*, *Arg264*, *His266*
2	17.2	60	*Pro89*, *Asp90*, *Asn91*, *Phe92*, *Val93*, *Phe94*, *Leu114*, *Ser117*, *Asp120*, *Val121*, *Lys124*, *Arg79*
3	16.7	91	Phe49, Asn50, Phe53, Arg123, Asp127, Cys129, Leu130, Phe135, Tyr161, Lys163, Lys164
4	15.5	57	Lys336, Thr337, Lys338, Arg339, Thr340, Gln342, Phe343, Asp345
5	13.8	97	Thr257, Asn258, Val260, Pro261, Tyr262, Trp346, Cys347, Pro348, Thr349
6	11.3	86	*Val177*, *Ser178*, *Thr180*, *Val181*, *Val182*, *Glu183*, *Pro184*, *Tyr185*, *Arg390*, *Glu393*, *Gln394*, *Phe395*, *Phe404*, *Lys176*
7	16.7	33	Gln176, Arg390, His393, Lys394, Leu397, *Leu333*, *Gln336*, *Asn337*, *Glu345*, *Trp346*, *Ile347*, *Pro348*, *Asn349*, *Asn350*
8	12.6	45	*Tyr36*, *Asp39*, *Ser40*, *Asp41*, *Leu42*, *Gln43*, *Ile358*, *Pro359*, *Arg369*

Table 5.9. Eight most important binding sites of FUOH on tubulin, the binding energies and populations (out of the total 1000 conformations) and the protein residues within 0.7 nm of FUOH. The α subunit residues are shown in normal and the β subunit residues in italic typeface.

The structure of the tubulin dimer was obtained as described previously (Section 5.4.1). In complex set, 10 FUOH molecules were placed at random locations around the protein. The simulations were carried out at 298 K and 1 bar. Four independent simulations, each lasting 50 ns were carried out for both sets.

(a) Tubulin without any FUOH molecules

Description. The tubulin dimer was simulated with its bound GTP and GDP molecules. Four independent simulations, each 50 ns long, were carried out.

Results. The two tubulin monomers remained bound to one another during each of the 50 ns simulations. Similarly the binding of GTP and GDP to the protein was also stable. However, there were some changes in the protein secondary structure. An overall decrease in the total number of residues involved in secondary structures was observed (see Fig. 5.15). When different types of secondary structures were examined separately, the number of residues forming α-helices showed a tendency to decrease while the number of residues forming β-sheets increased. This was not surprising because the GROMOS force field is known to have certain problems in preserving the secondary structures. It has a generic tendency to form β-sheets at the expense of other secondary structures and hence it is a "β-sheet friendly" force field [183]. The reason for choosing GROMOS force field despite this known problem

was that, it is the only force field for which the parameters for all the three types of molecules (protein, GTP/GDP and FUOH) were available. Further, other force fields also have similar problems in preserving the secondary structures [183].

(b) Tubulin with 10 FUOH molecules

Description. In this set, the tubulin dimer was simulated with ten FUOH molecules placed randomly around it. Four copies of the simulations were performed, each lasting 50 ns. For better sampling, the FUOH molecules were placed at different locations in each simulation.

Results. In all the simulations it was observed that FUOH binds to different sites with no single preferential binding site. It was also observed in docking simulations that FUOH has many equally important binding sites. Both these computational results are consistent with ITC experiments which indicated that FUOH binds at up to 9 sites [120].

Once bound to the sites, FUOH molecules remained bound there for the rest of the simulations. This can be seen in Fig. 5.16 in which the minimum distance between FUOH and tubulin residues is plotted for one simulation. At these sites FUOH formed a number of hydrogen bonds. On average, each FUOH molecule formed about six hydrogen bonds at the protein surface (see Fig. 5.16). This number is higher than the average number of hydrogen bonds FUOH formed with *taq* pol. The reason for this could be that tubulin has many flexible loops between secondary structures which could adjust easily to bind the nanoparticle. Further, when FUOH binds at the intradimer interface, the contact area with the protein is larger and hence FUOH makes contact with more residues there.

Experiments indicated a decrease in the total number of α-helical residues and an increase in the beta sheet residues when FUOH concentration was increased up to 100 mg/L [120]. To check this, the secondary structure of the protein was analyzed in the presence of FUOH. As in protein-only simulations, there was an overall tendency for a decrease in the number of α-helical residues and an increase in the β-sheet residues. This change is typical to GROMOS 53A6 force field [183] and was also observed in the previous set of simulations in which no FUOH molecules were included. When the two sets of simulations were compared (Fig. 5.15), the number

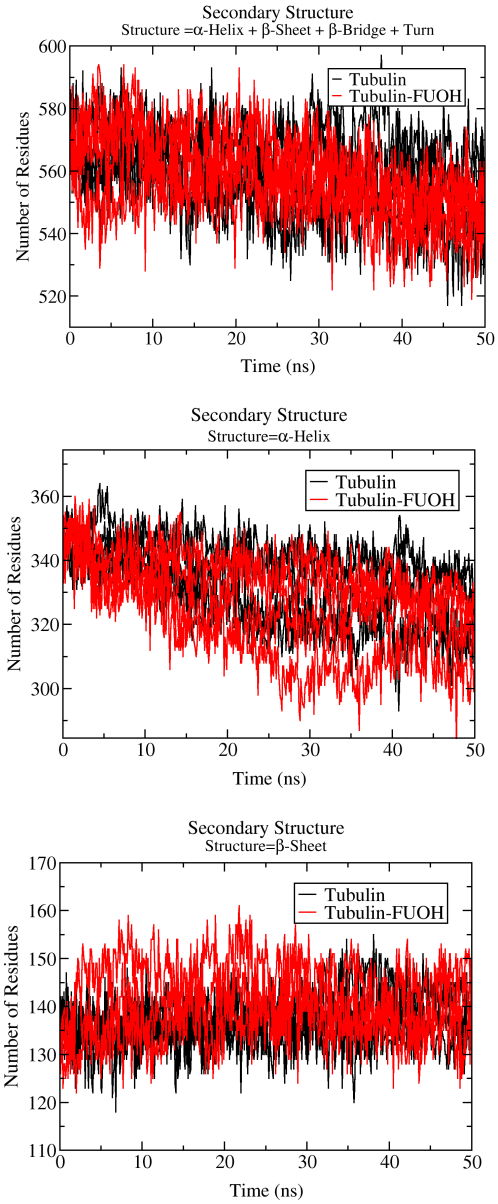


Figure 5.15. Secondary structures formed by tubulin with and without FUOH molecules. There is an overall β-sheet friendly tendency inherent to the force field. However, if the two sets of simulations are compared, the number of α-helical residues is slightly smaller and the number of β-sheet residues is slightly larger for the tubulin-FUOH system.

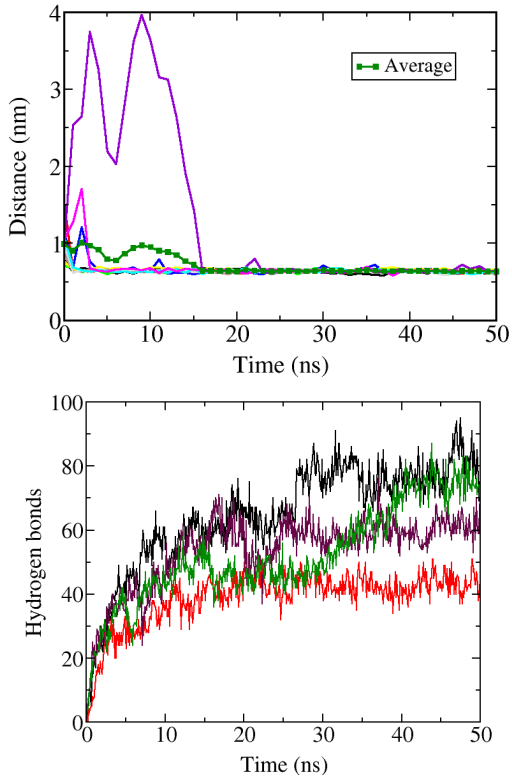


Figure 5.16. Top: Tubulin-FUOH minimum distance as a function of time for 10 FUOH molecules in an MD simulation. The average over 10 molecules is also shown. Bottom: The total number of hydrogen bonds that 10 FUOH molecules make in each simulation as a function of time.

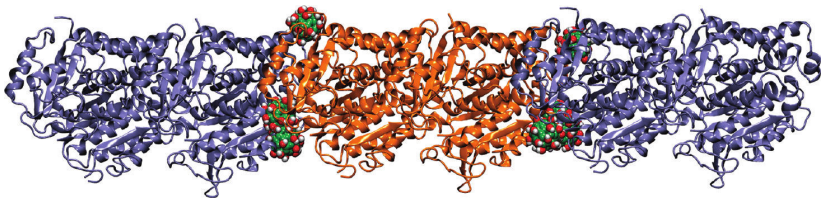


Figure 5.17. A tubulin dimer with its two longitudinal neighbors. The binding sites of FUOH at the longitudinal interface are also shown. Binding of FUOH at these sites would affect the dimer-dimer contact and hence the microtubule self-assembly.

of α-helical residues was slightly smaller (1.8 ± 1.1)% and the number of β-sheet residues was slightly larger (1.9±1.7)% in the presence of FUOH. Although this is consistent with the experimental observations, the differences are very small. Such changes in secondary structures of proteins due to nanoparticles have been reported in many studies [130], but the exact reason is not very well understood.

It was shown that FUOH can bind dNTPs with binding energies greater than 10 kJ/mol [109]. It is thus possible that FUOH molecules bind the GTP/GDP (which are very similar to dNTPs except for one hydroxyl group) in the E/N sites of tubulin. If FUOH binds to the E-site, it would affect the GDP/GTP exchange which plays an important role in the microtubule self-assembly. However, neither in the docking nor in the MD simulations these sites were found to be favorable sites for FUOH. It is not surprising that FUOH does not bind to the N-site because it is buried inside in the dimer. However, the E-site is exposed in the dimer and hence FUOH can bind there.

5.4.3 Possible mechanism

Based on the docking and MD simulation results, the following is a possible mechanism by which FUOH may affect the microtubule dynamics. The tubulin dimer makes a number of contacts with its neighbors when incorporated into a microtubule. These include longitudinal contacts with neighbors belonging to the same protofilaments (α-β interdimer contact) and lateral contacts with neighbors belonging to adjacent protofilaments (α-α and β-β contacts). If FUOH molecules bind to tubulin at these areas, they could affect the longitudinal and lateral contacts and hence the microtubule dynamics.

The length distribution and dynamics of microtubules are related to each other as mentioned in Section 2.4. But, can we say anything about the effects of FUOH on

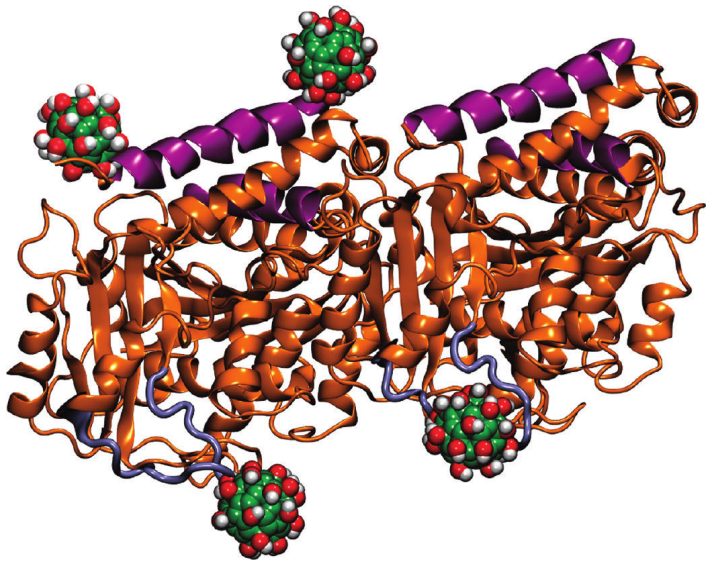


Figure 5.18. The interaction between the M-loop of one dimer (blue) and the helices H5 and H12 (purple) in a neighboring dimer contribute most to the interdimer lateral contacts in zinc sheets [26]. FUOH was observed to make contacts with these structures in the MD simulations.

microtubule dynamics (or microtubule length distribution) from our studies? The reason behind such a question is the fact that the structure of the tubulin dimer used in this study was obtained from X-ray crystallography of zinc-induced sheets [26], and not from microtubules themselves. In zinc-induced sheets the protofilaments are arranged in an anti-parallel fashion, unlike in microtubules in which they are parallel [24]. However, the longitudinal contacts between dimers in zinc sheets are similar to those in microtubules [26]. In Fig. 5.17, a tubulin dimer with its two longitudinal neighbors is shown. The binding sites of FUOH at these interfaces are also shown. Clearly, binding of FUOH at these sites would affect the dimer-dimer interactions and hence tubulin insertion into microtubules.

How about the lateral contacts? Contacts between the M-loop of one tubulin and H5 and H12 helices of an adjacent tubulin is a major contributor to the lateral interaction between tubulins in zinc sheets [26]. The M-loop is also involved in the lateral contacts in microtubules [26]. In Fig. 5.18 the binding of FUOH to M-loops and the two helices (as seen in the MD simulations) is shown. These bindings would obviously affect the lateral contact between tubulins in a microtubule.

Thus, we showed using our computational methods that multiple FUOH molecules bind to tubulin dimers forming a number of hydrogen bonds in agreement with indi-

cations from the ITC measurements. The secondary structure of tubulin is marginally affected by FUOH binding, in qualitative agreement with the experiments. The inhibition of microtubule assembly is likely due to the FUOH binding at lateral and longitudinal contact areas of tubulin dimers.

5.5 Ubiquitin-nanoparticle interactions

The interaction of fullerene (FUL) and fullerol (FUOH) with ubiquitin is described in this section. Both docking and molecular dynamics approaches were utilised. The motivation for this study is described in Section 3.4.

5.5.1 Docking

Description. FUL and FUOH nanoparticles were docked on to a ubiquitin molecule. The structure of ubiquitin was obtained from Protein Data Bank (PDB ID: 1UBQ [203]). Fifty docking simulations with 10 trials in each simulation were carried out for both FUL and FUOH.

Results. The docked conformations were clustered according to their mutual root mean square deviations. For FUL, only one major binding site was predicted (shown on the left side in Fig. 5.19). This site is located near the bend formed by residues Ala46 and Gly47. It has a binding energy of 29.7 kJ/mol and contributed about 90% to all the docked conformations.

For FUOH, two binding sites were observed (shown on the right side in Fig. 5.19). First site was same as the FUL binding site. It contributed about 57% of docked conformations with an energy of 34.3 kJ/mol. The residue Tyr59 is located in the vicinity of this binding site. According to fluorescence quenching experiments, FUOH molecules bind to the protein in the proximity of Tyr59 [125]. The second site contributed 34% of the conformations and had a binding energy of 33.9 kJ/mol. This site is located near the C-terminal tail of the protein. Residues which make contact with FUOH at this site are the terminal residues 72–76.

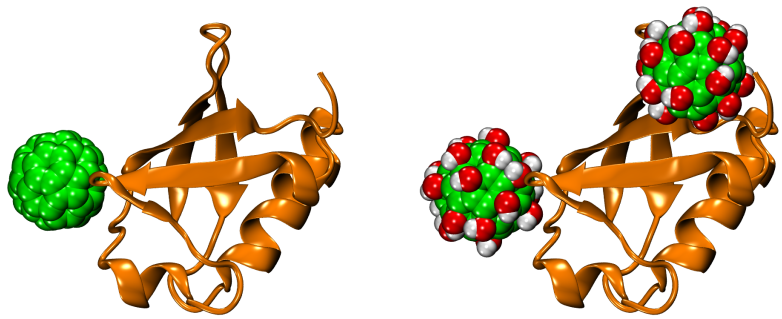


Figure 5.19. Binding sites of FUL and FUOH on ubiquitin molecule as predicted by docking simulations.

5.5.2 Molecular Dynamics

The systems simulated were (a) one ubiquitin molecule without any nanoparticles, (b) one ubiquitin with multiple nanoparticles distributed around it, and (c) one ubiquitin with multiple nanoparticles placed at one side of the protein.

(a) Ubiquitin without any nanoparticles

Description. In this set, the protein was simulated with no nanoparticles. As usual, this enabled us to compare the effect of nanoparticles on the protein properties. Six simulations, each lasting 100 ns, were carried out at 298 K.

Results. Ubiquitin is a small but very stable protein. The stability of the protein was reflected in the MD simulations. Overall, the structure remained unchanged with the exception of the long tail in the C-terminus which wiggled during the simulation. The rms deviation from the crystal structure was very small (average about 0.20 ± 0.03 nm) even with the inclusion of the fluctuating C-terminus. No significant change in the secondary structure was observed in the DSSP analysis (see Fig. 5.22).

(b) Multiple nanoparticles distributed around ubiquitin

Description. Ubiquitin was simulated with multiple nanoparticles distributed around it. The number of nanoparticles was either 5 or 13, and the nanoparticles were either

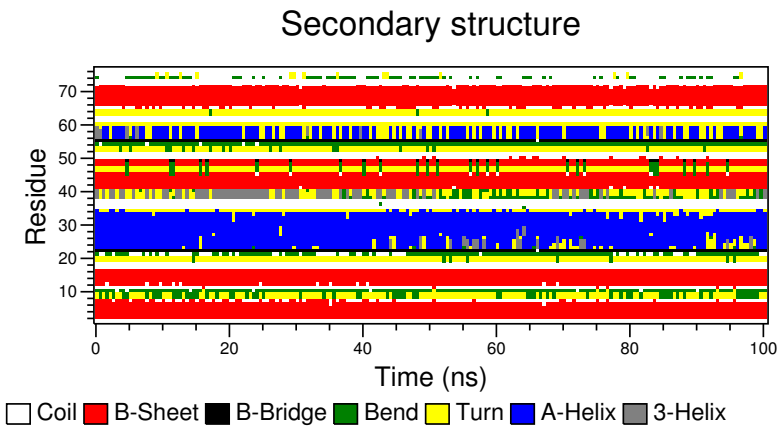


Figure 5.20. Secondary structures formed by ubiquitin residues in one of the simulations. The secondary structure analysis was performed using the program DSSP [199].

FUL or FUOH molecules. For each set, six independent simulations were performed, each lasting 100 ns.

Results. As the simulations were started, the nanoparticles were bound to the protein at various locations. There were two locations where nanoparticles (especially FUL) preferentially bound. One of these is near the C-terminal tail (upper right side in Fig. 5.21) and the second one near the turn formed by residues Ala46 and Gly47 (lower left side). In docking studies, these two sites were predicted as binding sites for FUOH, and one of them (lower left) for FUL. After binding to the protein surface, the nanoparticles showed a tendency to form aggregates. This tendency was more apparent for FUOH in comparison to FUL (See Fig. 5.21).

Upon nanoparticle binding, no overall change in the structure of the protein was observed. The secondary structure was well preserved just like in the protein-only simulations (see Fig. 5.22). However, the rms deviations of the protein coordinates was slightly smaller than the corresponding values in the protein-only simulations. This was likely due to binding of nanoparticles to the long C-terminal tail, which would reduce its fluctuations. The rms fluctuations of the protein residues were calculated, and the tail residues indeed showed lower fluctuations in the simulations where nanoparticles were present. The reduction in rms deviation (and rmsf of tail residues) was larger when 13 nanoparticles were distributed in comparison to the simulations with 5 nanoparticle. This is expected because with a larger number of nanoparticles, more nanoparticles bind to the C-terminal tail. The average number

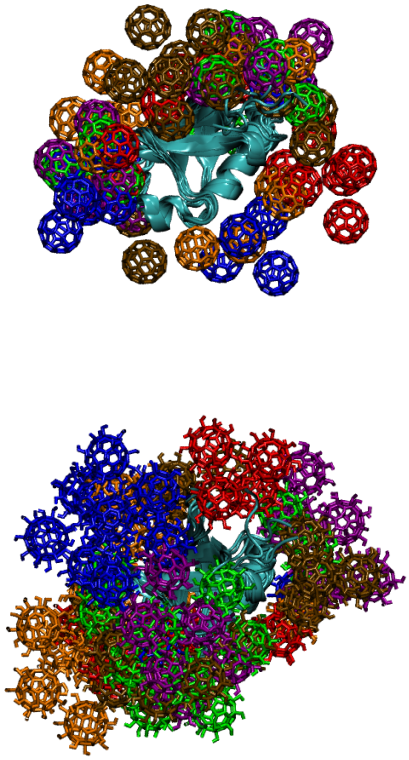


Figure 5.21. Snapshots of the structures formed by ubiquitin and 13 FUL (top) or 13 FUOH (bottom) at the end of 100 ns simulations. The structures of the protein from independent simulations are superimposed and the nanoparticles from independent simulations are colored differently. FUL seemed to have strong preferential locations whereas FUOH showed a high tendency to cluster on the protein surface.

of residues involved in secondary structures and the average rms deviations are listed for each set in Table 5.10.

(c) Multiple nanoparticles placed at one side of ubiquitin

Description. The simulations were performed by placing 13 FUL or FUOH nanoparticles at one side of the protein. In this set, the nanoparticles were able to cluster before binding to the protein. The simulations were carried out to understand how the clustering of nanoparticles affects the interaction with the protein. For FUL an additional set of simulation was carried out. Since fullerenes are known to form an icosahedron structure, 13 FUL molecules were arranged in the respective configuration at the beginning of these simulations. For each set, six 100 ns long independent simulations were performed.

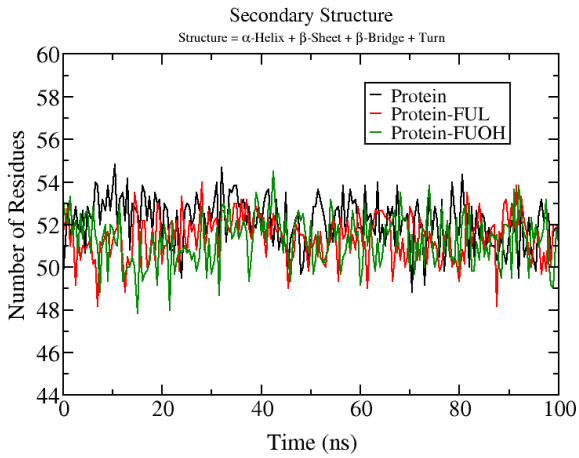


Figure 5.22. The average number of ubiquitin residues forming secondary structures in the absence and in the presence of 13 FUL or 13 FUOH molecules. The nanoparticles, if present, were distributed around the protein at the beginning of the simulations.

System	# Secondary structure	rms deviation (nm)	System description
Protein	52.1±0.5	0.20±0.03	Protein without any nanoparticle
Protein-FUL-n13	51.4±1.2	0.15±0.01	13 FUL distributed around the protein
Protein-FUOH-n13	51.3±1.0	0.16±0.02	13 FUOH distributed around the protein
Protein-FUL-n5	52.2±0.9	0.19±0.03	5 FUL distributed around the protein
Protein-FUOH-n5	52.1±0.8	0.17±0.03	5 FUOH distributed around the protein
Protein-FUL-side-n13	52.1±1.0	0.19±0.04	13 FUL placed at one side of the protein
Protein-FUOH-side-n13	52.0±0.5	0.20±0.04	13 FUOH placed at one side of the protein
Protein-FUL-side-icosa	52.5±0.5	0.20±0.02	13 FUL placed in an icosahedron configuration at one side of the protein

Table 5.10. Average number of ubiquitin residues involved in secondary structures and the rms deviation of the protein coordinates for different systems simulated. The standard deviation of the run-averages are also shown.

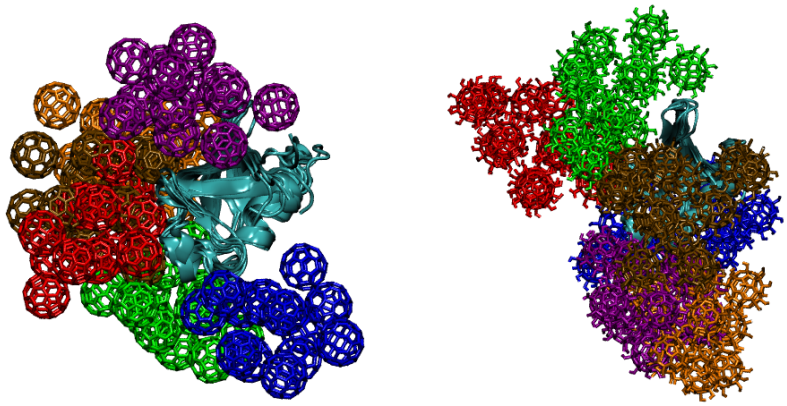


Figure 5.23. Snapshots of structures formed by ubiquitin and 13 FUL (left) or 13 FUOH (right) at the end of 100 ns simulations. The nanoparticles were placed together at one side of the protein (instead of distributing around the protein).

Results. In these simulations, the nanoparticles first formed aggregates before binding to the protein. The structures formed after 100 ns simulations are shown in Fig. 5.23. The apparent difference with the previous set of simulations was that both FUL and FUOH tended to remain as aggregate even after binding to the protein. In some of the simulations, the binding of the nanoparticle aggregate to the protein was minimal (see Fig. 5.23). When FULs were simulated in the icosahedron configuration, the structure remained intact and binding with the protein did not seem to have any effect. The rms deviation of FUL coordinates was about 0.5 nm. Similar values of rms deviation were observed for icosahedron structure in water in the absence of any other molecules (see Section 5.1).

Discussion

We observed that both the nanoparticles studied here bind to the protein in all sets of simulations. However, the nature of the binding depends on the type of nanoparticle and how they were distributed at the beginning of the simulation. In general, if the nanoparticles were distributed around the protein in the beginning, the binding was more uniform. On the other hand, if they were placed together at one side of the protein, allowing them to form aggregates before binding to the protein, the nanoparticles formed aggregates which then were bound to the protein. However, the binding of aggregates was minimal in some simulations. It is possible that the difference be-

tween the binding modes may be a result of the inadequate sampling time. However, we saw in Section 5.1 that nanoparticles can aggregate on much shorter time scales. Thus it is possible that the nature of the nanoparticle cluster can play a significant role in the interaction with proteins. Many experimental studies have indeed shown that the interaction between nanoparticles and proteins depends on the size and shape of the nanoparticles [124].

When nanoparticles bind to the C-terminal tail, the rms fluctuation of tail residues were reduced. The formation of polyubiquitin is extremely important in the functioningof ubiquitin-proteasome systems. Polyubiquitins are formed by the attachment of the C-terminal residue of one ubiquitin molecule to Lys48 or Lys63 residue of another (see Section 2.5). Because of the crucial role of the C-terminal residue, the binding of FUL and FUOH to the C-terminal tail and subsequent reduction in the flexibility of the tail is likely to have an impact on the ubiquitin-proteasome system. This can be asserted only after experimental studies.

6. Conclusions

The theme of this Dissertation is computational modeling of the interaction between proteins and fullerene-based nanoparticles. The motivation for this study comes from the growing concerns about biological effects of engineered nanoparticles. Many experiments have indeed shown that the normal functionalities of proteins were reduced as a result of their interaction with nanoparticles. Here, molecular docking and molecular dynamics simulations were used to understand the mechanisms behind the decreased activities. This combination was very effective to find the binding sites of nanoparticles on proteins, and the changes in the structure and dynamics of proteins upon nanoparticle binding. Since the functioning of a protein depends on it three-dimensional structure, insights into the mechanism of inhibition were obtained. The study on protein–nanoparticle interactions was augmented by two complementary studies. The properties of small clusters of fullerene-based nanoparticles, and the effects of these nanoparticles on α-helical propensities of small peptides were investigated.

According to this work, two fullerene-based nanoparticles — fullerol and fullerene trimalonic acid derivative — cause major tertiary structure changes to *taq* DNA polymerase. The inhibition of the activity of *taq* polymerase upon nanoparticle binding was attributed to these tertiary structure changes. The results were consistent with the experimental findings on the differences between the inhibitory activities of two nanoparticles. In the second study, fullerol was shown to have a tendency to bind to the M-loop and H5 and H12 helices of the tubulin. The interaction between M-loop of one tubulin and the H5 and H12 helices of the neighboring tubulin is a major contributor to the tubulin–tubulin binding energy in microtubules. In the third study the binding and its effects on ubiquitin were examined. The binding of fullerene

and fullerol nanoparticles seemed to depend on the nature of the nanoparticle cluster and also on the surface modifications. The flexibility of the C-terminal tail — which takes part in polyubiquitin formation — was reduced upon nanoparticle binding, implying potential effects on the protein degradation system. The nature and stability of nanoparticle clusters were found to depend heavily on the surface moieties of the nanoparticles. The clusters were packed and stable for fullerene which has a hydrophobic surface. For fullerol and trimalonic acid nanoparticles, which have hydrophilic surface groups, the clusters were less strongly packed and were dynamic. Finally, the stabilizing effects of nanoparticles on some small α-helical peptides was observed.

A common finding from this work is that the binding of the fullerene-based nanoparticles to the proteins studied is not specific. However, there were some locations on the protein surfaces which were preferential sites for the nanoparticles. Shape complementarity may have a role in this regard. The surface functional groups of the nanoparticles also had a great impact on binding. This was particularly true if the surface groups were charged. Changes in the structure and dynamics were able to be captured within the time scales of the simulations.

For docking simulations the flexibility of a protein can be important in some cases. In the work presented in this Dissertation the protein side chains at the binding sites were modeled as flexible. But by doing so, no significant changes were observed in the binding energies or binding modes. Further, in publication [120], conformations generated by MD simulations were used to account for the conformational changes of the protein.

The time scales involving α-helix formation is below hundred nanoseconds whereas the unfolding kinetics are much faster (less than 15 ns, depending on the length) [204]. The study on the secondary structures of peptides would still benefit from extending each simulation up to 100 ns or more. Since different force fields have tendencies to form particular secondary structures, it will also be important to repeat the simulations with different force fields in order to make sure that the differences observed are not due to any force field bias.

Bibliography

[1] H. J. C. Berendsen, *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, Cambridge University Press, 2007.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell 5th ed.*, Garland Science, 2008.

[3] Cartoon by Mariana Ruiz Villarreal. The image has been released into the public domain by its author, who grants anyone to use this work for any purpose, without any conditions, unless such conditions are required by law.

[4] J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry 5th ed.*, W. H. Freeman and Company, New York, 2002.

[5] Cartoon by Yassine Mrabet. The image has been released into the public domain by its author, who grants anyone to use this work for any purpose, without any conditions, unless such conditions are required by law.

[6] E. Ahmad, G. Rabbani, N. Zaidi, M. A. Khan, A. Qadeer, M. Ishtikhar, S. Singh, R. H. Khan, *J. Biomol. Struct. Dyn.*, 2012, Electronic publication ahead of print.

[7] Reprinted by permission from Macmillan Publishers Ltd.

[8] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Springer, 2002.

[9] J. A. Tuszynski, *Molecular and Cellular Biophysics*, Chapman and Hall/CRC, 2008.

[10] C. N. Pace, J. M. Scholtz, *Biophys. J.*, 1998, **75**, 422 – 427.

[11] N. J. Greenfield, G. D. Fasman, *Biochemistry*, 1969, **8**, 4108–4116.

[12] Y. Li, S. Korolev, G. Waksman, *EMBO J.*, 1998, **17**, 7514.

[13] T. A. Steitz, *J. Biol. Chem.*, 1999, **274**, 17395–17398.

[14] S. Korolev, M. Nayal, W. M. Barnes, E. D. Cera, G. Waksman, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 9264–9268.

[15] Y. Kim, S. H. Eom, J. Wang, D.-S. Lee, S. W. Suh, T. A. Steitz, *Nature*, 1995, **376**, 612.

[16] K. R. Tindall, T. A. Kunkel, *Biochemistry*, 1988, **27**, 6008–6013.

[17] W. M. Barnes, *Gene*, 1992, **112**, 29.

[18] P. J. Rothwell, G. Waksman, *Adv. Protein Chem.*, 2005, **71**, 401–440.

[19] K. B. Mullis, *Sci. Am.*, 1990, **262**, 56–61.

[20] H. A. Erlich, *J. Clin. Immuno.*, 1989, **9**, 437–447.

[21] O. C. Rodriguez, A. W. Schaefer, C. A. Mandato, P. Forscher, B. W. M., C. M. Waterman-Storer, *Nature Cell Biology*, 2003, **5**, 559.

[22] A. Desai, T. J. Mitchison, *Annu. Rev. Cell Dev. Biol.*, 1997, **13**, 83–117.

[23] M. Caplow, J. Shanks, S. Breidenbach, R. L. Ruhlen, *J. Biol. Chem.*, 1988, **263**, 10943–10951.

[24] E. Nogales, S. G. Wolf, K. H. Downing, *Nature*, 1998, **391**, 199–203.

[25] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33 – 38.

[26] J. Lowe, H. Li, K. Downing, E. Nogales, *J. Mol. Biol.*, 2001, **313**, 1045 – 1057.

[27] Reprinted by permission from Macmillan Publishers Ltd: Nature Rev. Cancer 4, 253-265 (2004). Copyright (2004).

[28] T. Mitchison, M. Kirschner, *Nature*, 1984, **312**, 237.

[29] H.-W. Wang, E. Nogales, *Nature*, 2005, **435**, 911–915.

[30] L. Jameson, M. Caplow, *J. Biol. Chem*, 1980, **255**, 2284–2292.

[31] J. Howard, A. A. Hyman, *Nature*, 2003, **422**, 753–758.

[32] E. Nogales, H.-W. Wang, *Curr. Opin. Cell Biol.*, 2006, **18**, 179–184.

[33] *The 2004 Nobel Prize in Chemistry - Popular Information*. `http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2004/popular.html`.

[34] A. Hershko, A. Ciechanover, *Annu. Rev. Biochem.*, 1998, **67**, 425–479.

[35] Cartoon by Rogerdodd via Wikimedia Commons, licenced under Creative Commons Attribution Share Alike 3.0.

[36] D. Hoeller, I. Dikic, *Nature*, 2009, **458**, 438–444.

[37] D. M. Parkin, *Int. J. Cancer.*, 2006, **118**, 3030–3044.

[38] `http://ec.europa.eu/dgs/jrc/downloads/jrc_reference_report_201007_nanomaterials.pdf`.

[39] A. D. Maynard, R. J. Aitken, T. Butz, V. Colvin, K. Donaldson, G. Oberdörster, M. A. Philbert, J. Ryan, A. Seaton, V. Stone, S. S. Tinkle, L. Tran, N. J. Walker, D. B. Warheit, *Nature*, 2006, **444**, 267.

[40] `http://www.discovernano.northwestern.edu/whatis/index_html/sizematters_html`.

[41] T. Da Ros, M. Prato, *Chem. Commun.*, 1999, 663–669.

[42] C. M. Sayes, J. D. Fortner, W. Guo, D. Lyon, A. M. Boyd, K. D. Ausman, Y. J. Tao, B. Sitharaman, L. J. Wilson, J. B. Hughes, J. L. West, V. L. Colvin, *Nano Lett.*, 2004, **4**, 1881–1887.

[43] O. Salata, *J. Nanobiotechnology*, 2004, **2**, 3.

[44] R. A. Sperling, P. Rivera Gil, F. Zhang, M. Zanella, W. J. Parak, *Chem. Soc. Rev.*, 2008, **37**, 1896–1908.

[45] M. Ferrari, *Nat. Rev. Cancer*, 2005, **5**, 161.

[46] D. Mijatovic, J. Eijkel, A. Berg, *Lab on a Chip*, 2005, **5**, 492.

[47] P. Harris, E. Hernandez, B. Yakobson, *Am. J. Phys.*, 2004, **72**, 415.

[48] `http://www.nanotechproject.org/inventories/consumer/analysis_draft/`.

[49] H. C. Fischer, W. C. Chan, *Curr. Opin. Biotechnol.*, 2007, **18**, 565–571.

[50] R. Behra, H. Krug, *Nat. Nanotechnol.*, 2008, **3**, 253.

[51] G. Oberdörster, F. Ferin, B. E. Lehnert, *Environ. Health Perspect.*, 1994, **102**, 173–179.

[52] G. Oberdörster, Z. Sharp, V. Atudorei, A. Elder, R. Gelein, W. Kreyling, C. Cox, *Inhalation Toxicol.*, 2004, **16**, 437.

[53] S. Foley, C. Crowley, M. Smaihi, C. Bonfils, B. F. Erlanger, P. Seta, C. Larroque, *Biochem. Biophys. Res. Commun.*, 2002, **294**, 116–119.

[54] S. Bosi, T. D. Ros, S. Castellano, E. Banfi, M. Prato, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 1043–1045.

[55] H. Aoshima, K. Kokubo, S. Shirakawa, M. Ito, S. Yamana, T. Oshima, *Biocontrol Sci.*, 2009, **14**, 69–72.

[56] D. Y. Lyon, L. K. Adams, J. C. Falkner, P. J. J. Alvarez, *Env. Sci. Tech.*, 2006, **40**, 4360–4366.

[57] S. Nakamura, T. Mashino, *Journal of Physics: Conference Series*, 2009, **159**, 012003.

[58] N. Tsao, T.-Y. Luh, C.-K. Chou, T.-Y. Chang, J.-J. Wu, C.-C. Liu, H.-Y. Lei, *J. Antimicrob. Chemother.*, 2002, **49**, 641–649.

[59] C.-X. Du, H.-R. Xiong, H. Ji, Q. Liu, H. Xiao, Z.-Q. Yang, *Sci. Res. Essays*, 2012, **7**, 705–711.

[60] S. H. Friedman, D. L. DeCamp, R. P. Sijbesma, G. Srdanov, F. Wudl, G. L. Kenyon, *J. Am. Chem. Soc.*, 1993, **115**, 6506.

[61] R. Bakry, R. M. Vallant, M. N. ul Haq, M. Rainer, Z. Szabo, C. W. Huck, G. K. Bonn, *Int. J. Nanomedicine*, 2007, **2**, 639–649.

[62] Y. Tabata, Y. Murakami, Y. Ikada, *Fullerene Sci. Technol.*, 1997, **5**, 989.

[63] N. Gharbi, M. Pressac, M. Hadchouel, H. Szwarc, S. R. Wilson, F. Moussa, *Nano Lett.*, 2005, **5**, 2578–2585.

[64] I. C. Wang, L. A. Tai, D. D. Lee, P. P. Kanakamma, C. K.-F. Shen, T.-Y. Luh, C. H. Cheng, K. C. Hwang, *J. Med. Chem.*, 1999, **42**, 4614–4620.

[65] H. Yamawaki, N. Iwai, *Am. J. Physiol. Cell Physiol.*, June 2006, **290**, C1495–C1502.

[66] S. Liu, H. Liu, Z. Yin, K. Guo, X. Gao, *J. Biomater. Nanobiotechnol.*, 2012, **3**, 385–390.

[67] E. Salonen, S. Lin, M. L. Reid, M. Allegood, X. Wang, A. M. Rao, I. Vattulainen, P. C. Ke, *Small*, 2008, **4**, 1986.

[68] E. Oberdörster, *Environ. Health Perspect.*, 2004, **112**, 1058–1062.

[69] W. R. Sanhai, J. H. Sakamoto, R. Canady, M. Ferrari, *Nat. Nanotechnol.*, 2008, **3**, 242–244.

[70] R. Taylor, *Lecture notes on fullerene chemistry*, Imperial college press, 1999.

[71] P. R. Buseck, S. J. Tsipursky, R. Hettich, *Science*, 1992, **257**, 215.

[72] J. E. Fischer, P. A. Heiney, A. B. Smith, *Acc. Chem. Res.*, 1992, **25**, 112–118.

[73] R. Taylor (Ed.), *The Chemistry of Fullerenes*, World Scientific, 1995.

[74] A. Dhawan, J. S. Taurozzi, A. K. Pandey, W. Shan, S. M. Miller, S. A. Hashsham, V. V. Tarabara, *Env. Sci. Tech.*, 2006, **40**, 7394–7401.

[75] R. S. Ruoff, D. S. Tse, R. Malhotra, D. C. Lorents, *J. Phys. Chem.*, 1993, **97**, 3379–3383.

[76] J. L. Atwood, G. A. Koutsantonis, C. L. Raston, *Nature*, 1994, **368**, 229.

[77] R. V. Bensasson, E. Bienvenue, M. Dellinger, S. Leach, P. Seta, *J. Phys. Chem.*, 1994, **98**, 3492–3500.

[78] H. Hungerbuehler, D. M. Guldi, K. D. Asmus, *J. Am. Chem. Soc.*, 1993, **115**, 3386–3387.

[79] Y. N. Yamakoshi, T. Yagami, K. Fukuhara, S. Sueyoshi, N. Miyata, *J. Chem. Soc., Chem. Commun.*, 1994, 517–518.

[80] P. C. Ke, R. Qiao, *J. Phys. Condens. Matt.*, 2007, **19**, 373101.

[81] M. Terashima, S. Nagao, *Chem. Lett.*, 2007, **36**, 302–303.

[82] K. L. Chen, M. Elimelech, *J. Colloid Interface Sci.*, 2007, **309**, 126–134.

[83] B. Espinasse, E. M. Hotze, M. R. Wiesner, *Env. Sci. Tech.*, 2007, **41**, 7396–7402.

[84] I. Lamparth, A. Hirsch, *J. Chem. Soc., Chem. Commun.*, 1994, 1727–1728.

[85] Y. Z. An, J. L. Anderson, Y. Rubin, *J. Org. Chem.*, 1993, **58**, 4799–4801.

[86] P. A. Heiney, J. E. Fischer, A. R. McGhie, W. J. Romanow, A. M. Denenstein, J. P. McCauley Jr., A. B. Smith, D. E. Cox, *Phys. Rev. Lett.*, 1991, **66**, 2911–2914.

[87] T. P. Martin, U. Näher, H. Schaber, U. Zimmermann, *Phys. Rev. Lett.*, 1993, **70**, 3079–3082.

[88] L. Y. Chiang, J. W. Swirczewski, C. S. Hsu, S. K. Chowdhury, S. Cameron, K. Creegan, *J. Chem. Soc., Chem. Commun.*, 1992, 1791–1793.

[89] J. Li, A. Takeuchi, M. Ozawa, X. Li, K. Saigo, K. Kitazawa, *J. Chem. Soc., Chem. Commun.*, 1993, 1784–1785.

[90] S. Wang, P. He, J. Zhang, H. Jiang, S. Zhu, *Synth. Commun.*, 2005, **35**, 1803–1808.

[91] S. Nakamura, K. Miyazawa, T. Mashino, *Preparation and Characterization of Fullerene Derivatives and Their Nanowhiskers*, Pan Stanford Publishing Ltd., 2011.

[92] R. Chen, T. A. Ratnikova, M. B. Stone, S. Lin, M. Lard, G. Huang, J. S. Hudson, P. C. Ke, *Small*, 2010, **6**, 612–617.

[93] Y. Tabata, Y. Ikada, *Pure Appl. Chem.*, 1999, **71**, 2047–2053.

[94] B.-X. Chen, S. R. Wilson, M. Das, D. J. Coughlin, B. F. Erlanger, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 10809–10813.

[95] Y.-L. Huang, C. K.-F. Shen, T.-Y. Luh, H. C. Yang, K. C. Hwang, C.-K. Chou, *Eur. J. Biochem.*, 1998, **254**, 38–43.

[96] S. Zhu, E. Oberdörster, M. L. Haasch, *Mar. Envi. Res.*, 2006, **62**, S5–S9.

[97] C. M. Sayes, A. M. Gobin, K. D. Ausman, J. Mendez, J. L. West, V. L. Colvin, *Biomaterials*, 2005, **26**, 7587–7595.

[98] J. Kamat, T. Devasagayam, K. Priyadarsini, H. Mohan, *Toxicology*, 2000, **155**, 55–61.

[99] J. Fang, D. Y. Lyon, M. R. Wiesner, J. Dong, Alvarez, *Env. Sci. Tech.*, 2007, **41**, 2636–2642.

[100] R. Qiao, A. P. Roberts, A. S. Mount, S. J. Klaine, P. C. Ke, *Nano Lett.*, 2007, **7**, 614–619.

[101] J. Wong-Ekkabut, S. Baoukina, W. Triampo, I.-M. Tang, D. P. Tieleman, L. Monticelli, *Nat. Nanotechnol.*, 2008, **3**, 363.

[102] C. F. Lopez, S. O. Nielsen, P. B. Moore, M. L. Klein, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 4431–4434.

[103] A. S. Boutorine, M. Takasugi, C. Helene, H. Tokuyama, H. Isobe, E. Nakamura, *Angew. Chem. Int. Ed.*, 1995, **33**, 2462–2465.

[104] Y. N. Yamakoshi, T. Yagami, S. Sueyoshi, N. Miyata, *J. Org. Chem.*, 1996, **61**, 7236–7237.

[105] A. Kumar, M. V. Rao, S. K. Menon, *Tetrahedron Lett.*, 2009, **50**, 6526–6530.

[106] S. Samal, K. E. Geckeler, *Macromol. Biosci.*, 2001, **1**, 329–331.

[107] X. Zhao, A. Striolo, P. T. Cummings, *Biophys. J.*, 2005, **89**, 3856–3862.

[108] X. Zhao, *J. Phys. Chem. C*, 2008, **112**, 8898–8906.

[109] J. Shang, T. A. Ratnikova, S. Anttalainen, E. Salonen, P. C. Ke, H. T. Knap, *Nanotechnology*, 2009, **20**, 415101.

[110] C. Ratna Prabha, R. Patel, C. N. Murthy, *Fullerenes, Nanotubes and Carbon Nanostruct.*, 2005, **12**, 405–412.

[111] S.-T. Yang, H. Wang, L. Guo, Y. Gao, Y. Liu, A. Cao, *Nanotechnology*, 2008, **19**, 395101.

[112] B. Belgorodsky, L. Fadeev, J. Kolsenik, M. Gozin, *ChemBioChem*, 2006, **7**, 1783–1789.

[113] B. Belgorodsky, L. Fadeev, V. Ittah, H. Benyamini, S. Zelner, D. Huppert, A. B. Kotlyar, M. Gozin, *Bioconjugate Chemistry*, 2005, **16**, 1058–1062.

[114] T. Mashino, K. Shimotohno, N. Ikegami, D. Nishikawa, K. Okuda, K. Takahashi, S. Nakamura, M. Mochizuki, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 1107–1109.

[115] N. Iwata, T. Mukai, Y. N. Yamakoshi, S. Haraa, T. Yanase, M. Shoji, T. Endo, N. Miyata, *Fullerene Sci. Technol.*, 1998, **6**, 213–226.

[116] X. Meng, Z. Chen, B. Li, Y. Zhang, D. Zhao, X. Yang, *Chin. Sci. Bull.*, 2006, **51**, 2550.

[117] Z. Zhu, D. I. Schuster, M. E. Tuckerman, *Biochemistry*, 2003, **42**, 1326–1333.

[118] S. Kraszewski, M. Tarek, W. Treptow, C. Ramseyer, *ACS Nano*, 2010, **4**, 4158–4164.

[119] S. H. Friedman, P. S. Ganapathi, Y. Rubin, G. L. Kenyon, *J. Med. Chem.*, 1998, **41**, 2424–2429.

[120] T. A. Ratnikova, P. Nedumpully Govindan, E. Salonen, P. C. Ke, *ACS Nano*, 2011, **5**, 6306–6314.

[121] X. Meng, B. Li, Z. Chen, L. Yao, D. Zhao, X. Yang, M. He, Q. Yu, *J. Enzyme Inhib. Med. Chem.*, 2007, **22**, 293–296.

[122] X. Yang, Z. Chen, X. Meng, B. Li, X. Tan, *Chin. Sci. Bull.*, 2007, **52**, 1802–1806.

[123] R. Pieper, C. L. Gatlin, A. J. Makusky, P. S. Russo, C. R. Schatz, S. S. Miller, Q. Su, A. M. McGrath, M. A. Estock, P. P. Parmar, M. Zhao, S.-T. Huang, J. Zhou, F. Wang, R. Esquer-Blasco, N. L. Anderson, J. Taylor, S. Steiner, *Proteomics*, 2003, **3**, 1345–1364.

[124] I. Lynch, K. A. Dawson, *Nanotoday*, 2008, **3**, 40–47.

[125] P.-C. Ke, private communication (2012).

[126] M.-F. Zhang, Z.-Q. Xu, Y.-S. Ge, F.-L. Jiang, Y. Liu, *J. Photochem. Photobiol. B: Biol.*, 2012, **108**, 34–43.

[127] W. Norde, C. E. Giacomelli, *J. Biotechnol.*, 2000, **79**, 259–268.

[128] X.-C. Shen, X.-Y. Liou, L.-P. Ye, H. Liang, Z.-Y. Wang, *J. Colloid Interface Sci.*, 2007, **311**, 400–406.

[129] F. Bellezza, A. Cipiciani, M. A. Quotadamo, S. Cinelli, G. Onori, S. Tacchi, *Langmuir*, 2007, **23**, 13007–13012.

[130] L. Fei, S. Perrett, *Int. J. Mol. Sci.*, 2009, **10**, 646–655.

[131] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198–1229.

[132] H. Lin, D. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185–199, 10.1007/s00214-006-0143-z.

[133] A. R. Leach, *Molecular Modelling: Principles and Applications*, Pearson Education Limited, 2001.

[134] B. J. Alder, T. E. Wainwright, *J. Chem. Phys.*, 1957, **27**, 1208.

[135] J. A. McCammon, B. R. Gelin, M. Karplus, *Nature*, 1977, **267**, 585, First MD of protein macromolecule.

[136] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, B. Towles in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, ACM, New York, NY, USA, of *SC '09*, pp. 39:1–39:11. `http://doi.acm.org/10.1145/1654059.1654099`.

[137] J. W. Ponder, D. A. Case, *Adv. Protein Chem.*, 2003, **66**, 27–85.

[138] A. D. Mackerell, *J. Comput. Chem.*, 2004, **25**, 1584–1604.

[139] B. Hess, D. van der Spoel, E. Lindahl, et al., *GROMACS User Manual Version 4.5.4*.

[140] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

[141] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins*, 2006, **65**, 712–725.

[142] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.

[143] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

[144] A. D. Mackerell, M. Feig, C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

[145] A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

[146] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.

[147] W. L. Jorgensen, D. S. Maxwell, J. T. Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

[148] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

[149] C. Oostenbrink, A. Villa, A. E. Mark, W. F. Van Gunsteren, *J. Comput. Chem.*, 2004, **25**, 1656–1676.

[150] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, H. B. Yu, *Angew. Chem. Int. Ed.*, 2006, **45**, 4064–4092.

[151] A. D. MacKerell, *Protein Force Fields*, John Wiley & Sons, Ltd, 2002.

[152] R. Hockney, S. Goel, J. Eastwood, *J. Comput. Phys*, 1974, **14**, 148 – 158.

[153] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327 – 341.

[154] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463.

[155] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

[156] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[157] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys*, 1995, **102**, 5451–5459.

[158] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684.

[159] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[160] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*, World Scientific, 2004.

[161] S. NosÃ©, *Molecular Physics*, 1984, **52**, 255–268.

[162] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[163] M. Parrinello, A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182.

[164] S. Nose, M. Klein, *Mol. Phys.*, 1983, **50**, 1055–1076.

[165] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell., *Molecular Cell Biology*, W. H. Freeman, 2000.

[166] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, D. E. Shaw, *Annu. Rev. Biophys.*, 2012, **41**, 429.

[167] M. Karplus, J. A. McCammon, *Nat. Struct. Biol.*, 2002, **9**, 646–652.

[168] Y. Sugita, *Chem. Phys. Lett.*, 1999, **314**, 141–151.

[169] G. A. Voth (Ed.), *Coarse-Graining of Condensed Phase and Biomolecular Systems*, CRC Press, 2008.

[170] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S.-J. Marrink, *J. Chem. Theory Comput.*, 2008, **4**, 819–834.

[171] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, K. L. Schulten, *Structure*, 2009, **17**, 1295.

[172] T. Lengauer, M. Rarey, *Curr. Opin. Struct. Biol.*, 1996, **6**, 402 – 406.

[173] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat. Rev. Drug Discov.*, 2004, **3**, 935–949.

[174] A. Bonvin, *Curr. Opin. Struct. Biol.*, 2006, **16**, 194–200.

[175] N. Brooijmans, I. D. Kuntz, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 335–373.

[176] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639.

[177] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2787.

[178] R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell, *J. Comput. Chem.*, 2007, **28**, 1145–1152.

[179] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, *J. Am. Chem. Soc.*, 1984, **106**, 765–784.

[180] D. N. A. Boobbyer, P. J. Goodford, P. M. McWhinnie, R. C. Wade, *J. Med. Chem.*, 1989, **32**, 1083–1094.

[181] J. Gasteiger, M. Marsili, *Tetrahedron*, 1980, **36**, 3219 – 3228.

[182] R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell, *J. Comput. Chem.*, 2007, **28**, 1145.

[183] D. Matthes, B. L. de Groot, *Biophys. J.*, 2009, **97**, 599–608.

[184] L. A. Girifalco, *J. Phys. Chem.*, 1992, **96**, 858–861.

[185] J. Uusitalo, *Permeation of fullerene derivatives through a POPC lipid bilayer*, Special Assignment Tfy-105.5111, Aalto University, 2010.

[186] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

[187] H. Berendsen, J. Postma, W. van Gunsteren, J. Hermans, *Interaction models for water in relation to protein hydration*, Reidel Dordrecht, 1981.

[188] P. Nelson, *Biological Physics: Energy, Information, Life*, W. H. Freeman and Company, 2004.

[189] C. L. Dias, T. Ala-Nissila, M. Karttunen, I. Vattulainen, M. Grant, *Phys. Rev. Lett.*, 2008, **100**, 118101.

[190] D. Chandler, *Nature*, 2005, **437**, 640–647.

[191] C. Bystroff, D. Baker, *J. Mol. Biol.*, 1998, **281**, 565 – 577.

[192] W. L. Delano, *The PyMOL Molecular Graphics System*, `http://www.pymol.org`, 2002. `http://www.pymol.org`.

[193] P. Nygren, M. Lundqvist, K. Broo, B.-H. Jonsson, *Nano Lett.*, 2008, **8**, 1844–1852, PMID: 18540660.

[194] K. Ikeda, T. Okada, S. ichi Sawada, K. Akiyoshi, K. Matsuzaki, *FEBS Lett.*, 2006, **580**, 6587 – 6595.

[195] M. Karplus, D. L. Weaver, *Protein Sci.*, 1994, **3**, 650–668.

[196] W. M. Barnes, *Proc. Nat. Acad. Sci.*, 1994, **91**, 2216.

[197] R. Saiki, D. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. Horn, K. Mullis, H. Erlich, *Science*, 1988, **239**, 487–491.

[198] K. B. Mullis, F. A. Faloona, *Methods Enzymol.*, 1987, **155**, 335–350.

[199] W. Kabsch, C. Sander, *Biopolymers*, 1983, **22**, 2577.

[200] S. H. Eom, J. Wang, T. A. Steitz, *Nature*, 1996, **382**, 278.

[201] A. Šali, T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779–815.

[202] A. Fiser, R. K. G. Do, A. Sali, *Protein Sci.*, 2000, **9**, 1753–1773.

[203] S. Vijay-Kumar, C. E. Bugg, W. J. Cook, *Journal of Molecular Biology*, 1987, **194**, 531 – 544.

[204] C. L. Brooks, *J. Phys. Chem.*, 1996, **100**, 2546–2549.

Aalto-DD 12/2013

ISBN 978-952-60-4979-3
ISBN 978-952-60-4980-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Applied Physics
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS